AD-A083 931 DREXEL UNIV PHILADELPHIA PA DEPT OF MECHANICAL ENGIN--ETC F/6 14/2
GENERALIZED APPROACH TO NEW PROBLEMS IN ULTRASONIC INSPECTION (--ETC(U)
APR 80 J L ROSE, M J AVIOLI N68335-78-C-1041
UNCLASSIFIED NAEC-92-140 NL

1 OF 2
AD
A083931

LEVEL II

72

NAVAL AIR ENGINEERING CENTER

REPORT NAEC-92-140

LAKEHURST, N.J.
08733

GENERALIZED APPROACH TO NEW PROBLEMS
IN ULTRASONIC INSPECTION (GANPUI)

Handling & Servicing/Armament Division
Ground Support Equipment Department
Naval Air Engineering Center
Lakehurst, New Jersey 08733

2 APRIL 1980

APPROVED FOR PUBLIC RELEASE:
DISTRIBUTION UNLIMITED

Prepared for

Commander, Naval Air Systems Command

Washington, DC 20361

NAEC-92-140

GENERALIZED APPROACH TO NEW PROBLEMS
IN ULTRASONIC INSPECTION (GANPUI)

Prepared by: Joseph L. Rose
JOSEPH L. ROSE
Mechanical Engr. & Mechanics Dept.
Drexel University

Michael J. Avioli, Jr.
MICHAEL J. AVIOLI
Mechanical Engr. & Mechanics Dept.
Drexel University

Reviewed by: R C Deitrich
R. C. DEITRICH
Advanced Technology Section, Handling
and Servicing/Armament Division

Approved by: F E Evans
F. E. EVANS
Ground Support Equipment Superintendent

NOTICE

4/2/80 da

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER NAEC-92-140	2. GOVT ACCESSION NO. AD-A083 931	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) Generalized Approach to New Problems in Ultrasonic Inspection (GANPUI)		5. TYPE OF REPORT & PERIOD COVERED
		6. PERFORMING ORG. REPORT NUMBER
7. AUTHOR(s) Dr. Joseph L. Rose; Mr. Michael J. Avioli		8. CONTRACT OR GRANT NUMBER(s) N68335-78-C-1041 new
9. PERFORMING ORGANIZATION NAME AND ADDRESS Drexel University, Mechanical Engr. and Mechanics Dept., Philadelphia, PA 19104		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS
11. CONTROLLING OFFICE NAME AND ADDRESS Naval Air Systems Command, Air-340E, Washington, DC 20361		12. REPORT DATE 2 Apr 1980
		13. NUMBER OF PAGES 97
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office) Naval Air Engineering Center, Ground Support Equip. Dept. (Code 92724), Lakehurst, NJ 08733		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

12 90

Approved for Public Release; Distribution Unlimited

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Feature Vector, probability density function, cluster analysis, algorithms, adaptive learning, Fisher linear discriminant

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Generalized approach to new problems in ultrasonic inspection is detail... It is an operator-computer interactive scheme involving application of latest techniques in ultrasonic inspection, pattern recognition, and minicomputer technology.

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-LF-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

390 ...

TABLE OF CONTENTS

LIST OF TABLES

LIST OF ILLUSTRATIONS

LIST OF ILLUSTRATIONS (Continued)

This page left blank
intentionally.

I. INTRODUCTION

A. GANPUI is an acronym for "A Generalized Approach to New Problems in Ultrasonic Inspection". Conceptually, it is an operator-computer interactive scheme that involves the application of the latest techniques in ultrasonic inspection, pattern recognition, and minicomputer technology. The following paragraphs offer a brief overview of GANPUI. Detailed descriptions of GANPUI components will comprise the main body of the text. Sample applications of GANPUI are also included.

B. The entire procedure may be divided into three general categories: input, processing, and output.

1. The input consists of several sub-processes, the first of which is the acquisition of reliable ultrasonic waveforms. These waveforms are then digitized, that is, decomposed into discrete time sequences. The PDP 11/05e minicomputer, for example, accepts these sequences and performs certain mathematical operations on them, such as determining maximums and minimums. This process is known as feature extraction. Features that are useful in ultrasonic examination include center frequency, 6 dB down bandwidth, energy ratios over specified frequency intervals of a frequency spectra, etc. At this stage, the computer operator selects the particular features to be employed in algorithm development. Once the feature values have been computed, they are stored in a vector filing system. This completes the input stage of the system.

2. Processing utilizes several complex schemes or algorithms to search for innate groupings of feature data. These data values are ordered with respect to their effect in defining particular groupings. Upon completion of processing, the significant features are combined to develop a classification scheme.

3. GANPUI output assists the ultrasonic investigator by minimizing the data acquired in decision making. Algorithm development makes use of many techniques in learning network analysis and in pattern recognition, the goal being to establish some relationship between classification mode and a number of important ultrasonic signal features. Regression analysis is considered at various points of the algorithm development process. Such techniques as probability density function analysis, cluster analysis, minimum distance classification, adaptive learning polynomials, and a Fisher linear discriminant are currently being used in our algorithm development test system.

C. An essential element in the GANPUI program of study is associated with the utilization of good training data. Good test samples are required so that the computer can be trained to recognize certain patterns. Test samples are obviously required to evaluate the classification algorithms developed by GANPUI.

D. Several problems that are being studied by various ultrasonic research groups that make use of GANPUI concepts include composite material inspection, aircraft and space shuttle adhesive bond evaluation, the detection of stress

corrosion cracking in stainless steel piping for the nuclear industry, flaw growth propagation in the shipping and aircraft industries, and the early detection of breast cancer in the field of diagnostic ultrasound.

E. A flow chart of GANPUI is shown in Table 1.

Table 1. GANPUI Flow Chart

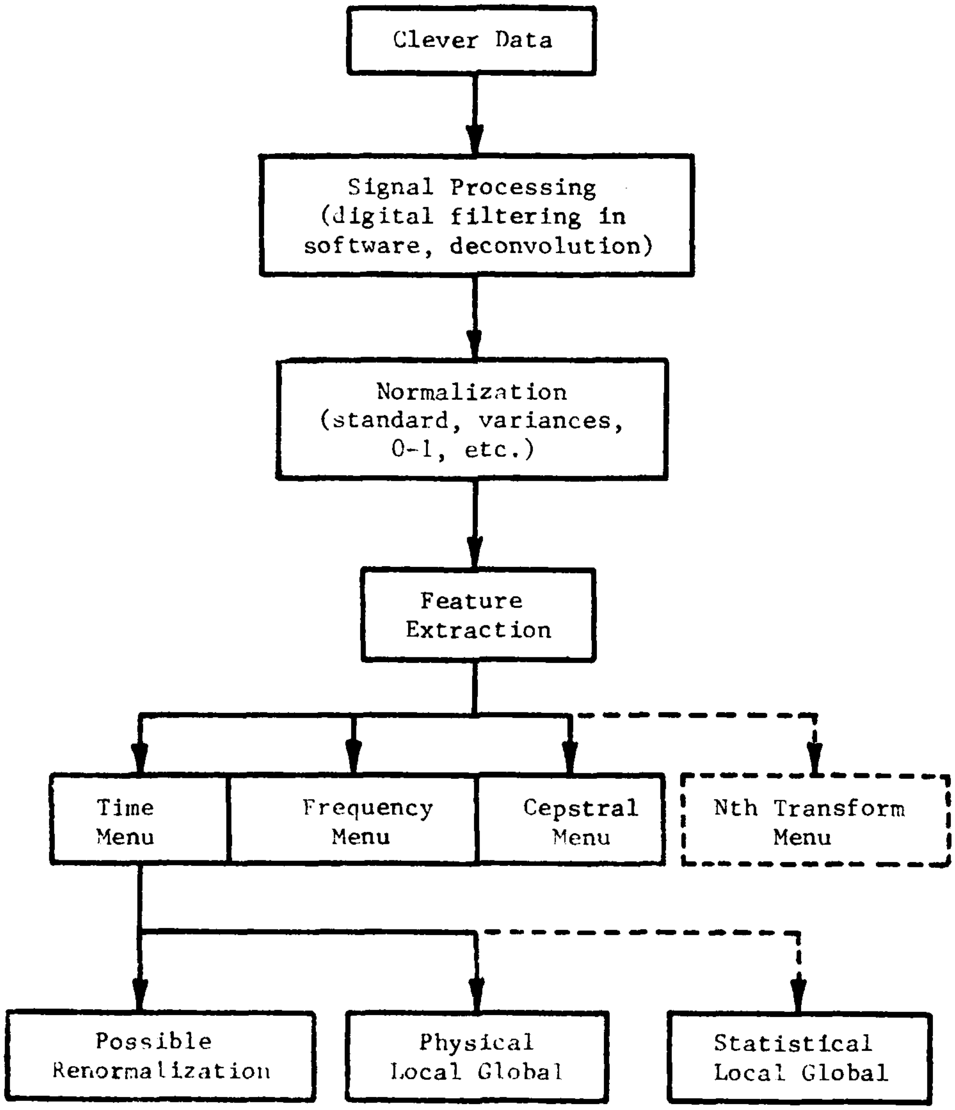
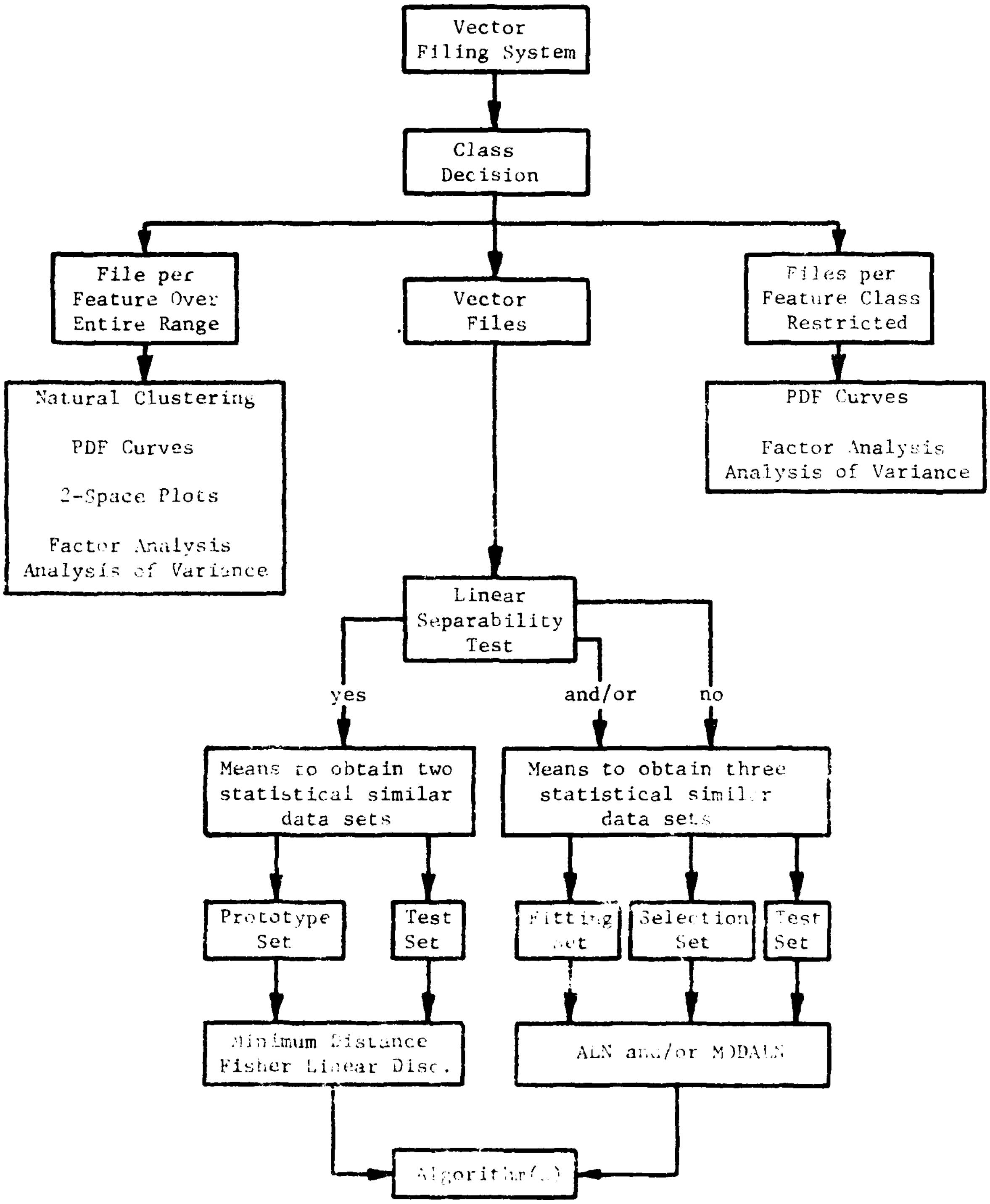


Table 1. GANPUI Flow Chart (Continued)



II. FOURIER SERIES AND TRANSFORM THEORY

A. FOURIER DECOMPOSITION. From a mathematical viewpoint, any arbitrary mathematical function could be resolved or decomposed into a finite sum of some other functions.

1. For example, if we were to examine the mathematical response of a system from a rectangularly shaped input function, we could, if mathematically convenient, compute the response from two rectangularly shaped functions as shown in Figure 1. In this particular example, the decomposition process into two rectangularly shaped functions does not provide us with any additional information or mathematical computational efficiency. The rectangular input pulse given above could, however, be decomposed into a sum of some other waveforms; say rectangular segments of fixed pulse duration of which a mathematical solution might be readily available in the literature or in a computer. The total solution could then be obtained by examining the contributions from each smaller segment and adding them together in linear fashion to obtain total solutions.

2. Let us consider now a different form for the input function. The pulse form shown below could be treated in a mathematical sense as a finite number of rectangularly shaped input pulse forms as illustrated in Figure 2. In the limiting process, as the pulse duration, δ, of the rectangularly shaped pulse decreases to some infinitesimal value, the response function or solution to some system could be identical.

3. A more useful function decomposition approach exists in a mathematical sense than those presented above in the rectangular segment approach. This approach is called frequency analysis or Fourier Series analysis. In this particular approach, a given functional shape is resolved into a finite number of sinusoidal waveforms. As an example, an arbitrary pulse form could be considered mathematically as the sum illustrated in Figure 3; the functions shown on the right are representing continuous wave or sinusoidal waveforms. As the number of waves being added together increases, the more accurate will be the comparison of the resulting waveforms with the initial waveform. As an example, if a rectangular pulse were to be resolved into one continuous wave term, the approximation would not be good. Two terms would be better. As shown in the example in Figure 4, considering the rectangular function as an odd function, adding both first and third harmonics, gives the approximation of the original function. In order to obtain the corners of the rectangular pulse, discontinuities being critical points and difficult points to approximate in a Fourier series sense, a fairly large number of terms, say [illegible] or more, would be required to approach the edges of the rectangular pulse

4. The Fourier Series of the periodic function is defined by

$$f(t) = \sum_{n=-\infty}^{\infty} a_n e^{jn2\pi \frac{t}{T}}$$

where f(t) is such that

$$f(t) = f(t + T); \quad T \equiv \text{period}$$

and

$$a_n = \frac{1}{T} \int_0^T f(t) e^{-jn2\pi \frac{t}{T}} dt$$

The a_n are called Fourier coefficients. Another common notation is

$$A_n = \text{Real } [a_n] = \frac{1}{T} \int_0^T f(t) \cos(n \frac{2\pi}{T} t) dt$$

$$B_n = \text{Im } [a_n] = \frac{1}{T} \int_0^T f(t) \sin(n \frac{2\pi}{T} t) dt$$

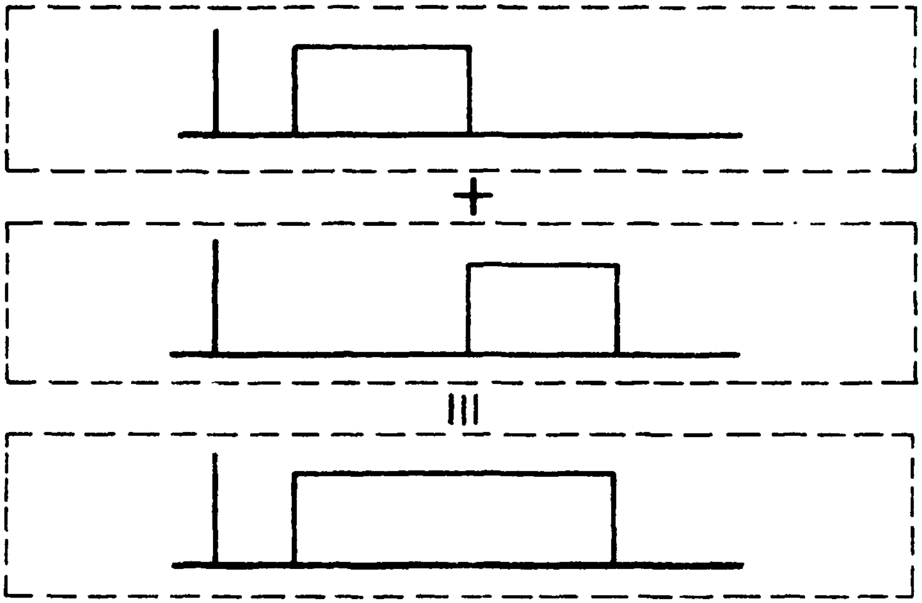


FIGURE 1. Rectangularly Shaped Functions

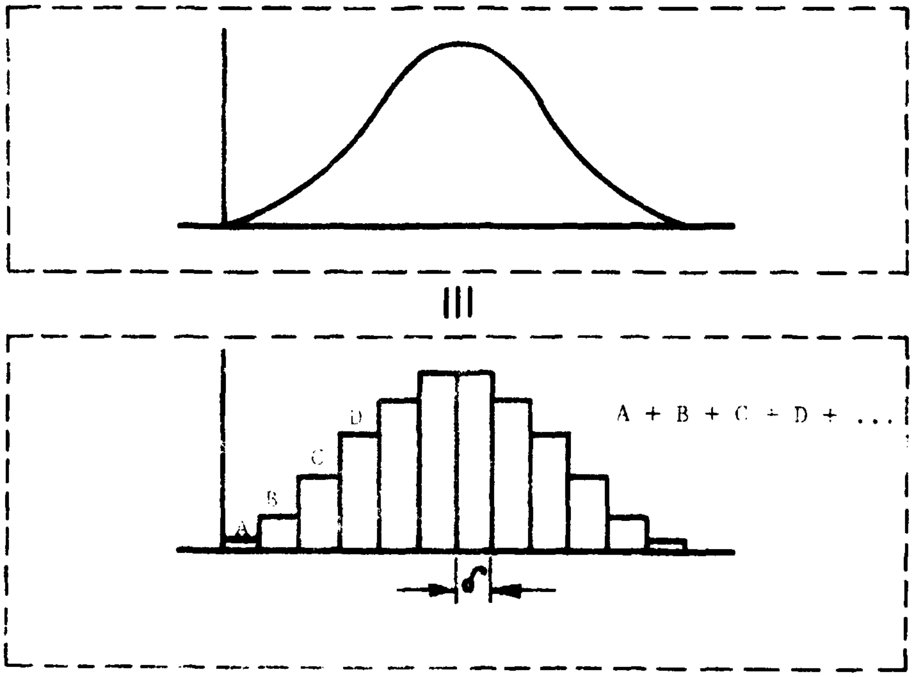


FIGURE 2. Pulse Form (Function Decomposition Concept)

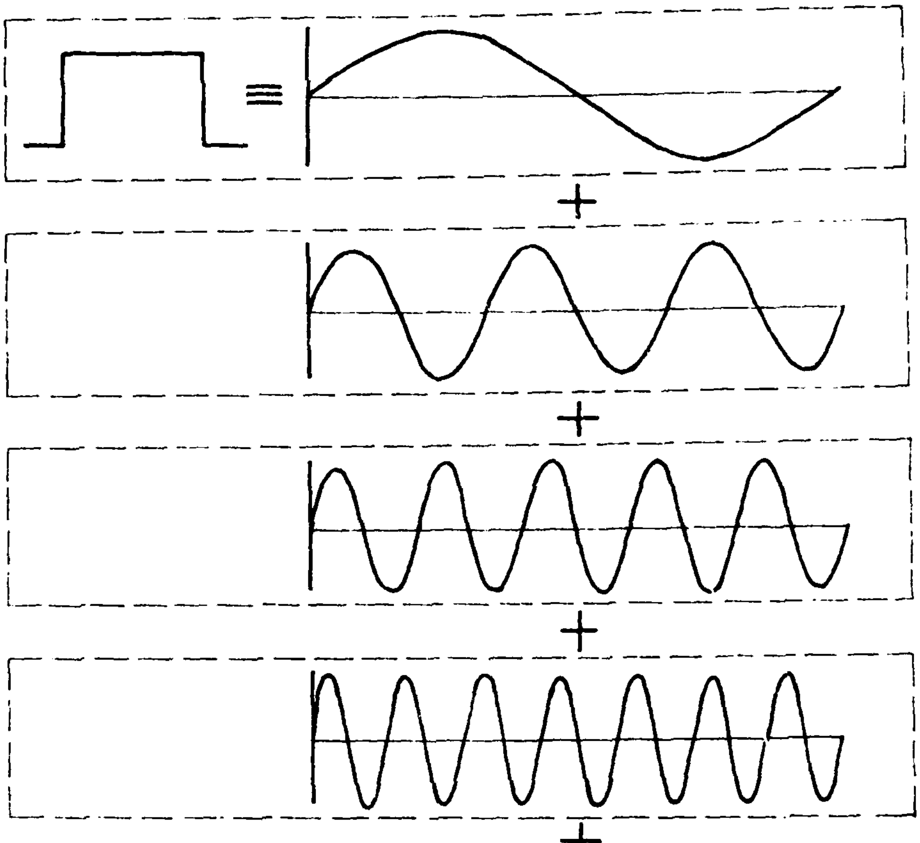


FIGURE 3. Arbitrary Pulse Wave Forms

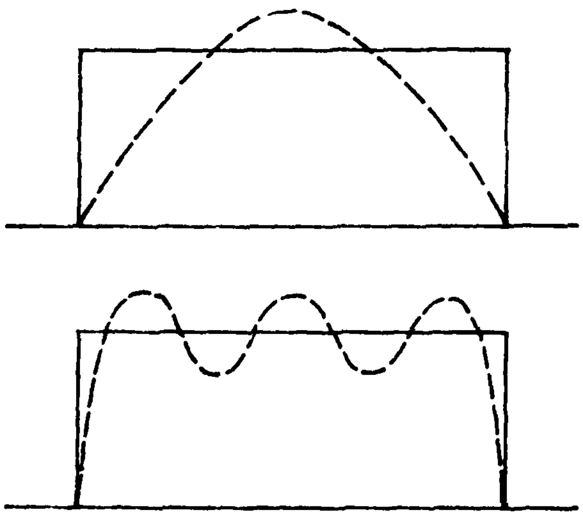


FIGURE 4. Fourier Series Approximation of a Rectangular Pulse

B. <u>FOURIER TRANSFORM THEORY</u>. Much of the pattern recognition work associated with ultrasonics, uses the Fourier transform of a time domain signal as a feature source. A feature is a parameter defined on a function. Consider the graphical representation of a function shown below in Figure 5.

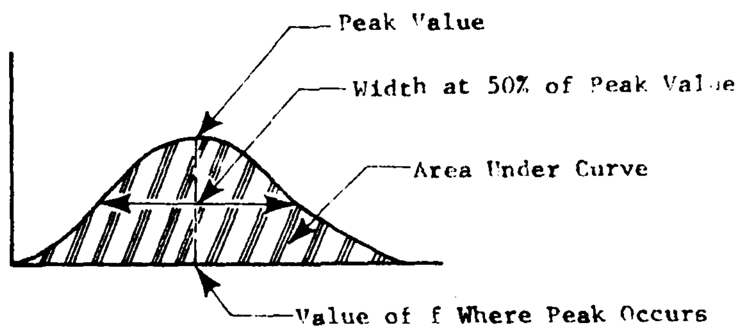


Figure 5. Example of Fourier Spectrum Parameterization

Three possible features are shown on the illustration. The Fourier transform will be reviewed in this section because of its importance in ultrasonic work.

By definition, the Fourier transform of a function of time f(t) is a function of angular frequency $F(\omega)$, given by the relationship

$$F(\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}\,dt$$

where $e^{-j\omega t} = \cos\omega t - j\sin\omega t \qquad (j = \sqrt{-1})$

$$\mathrm{Re}[F(\omega)] = \int_{-\infty}^{\infty} f(t)\cos\omega t\,dt \qquad \text{(real part)}$$

$$\mathrm{Im}[F(\omega)] = \int_{-\infty}^{\infty} -f(t)\sin\omega t\,dt \qquad \text{(imaginary part)}$$

Terms usually associated with the Fourier transform are Power Spectrum and Phase Angle. These are defined below.

$$\text{Power Spectrum} \equiv |F(\omega)|^2 = (\mathrm{Re}[F(\omega)])^2 + (\mathrm{Im}[F(\omega)])^2$$

$$\text{Phase Angle} \equiv \phi(\omega) = \tan^{-1}(\mathrm{Im}[F(\omega)]/\mathrm{Re}[F(\omega)])$$

An example of a time function along with its spectrum and phase angle is given in Figure 6 on the following page.

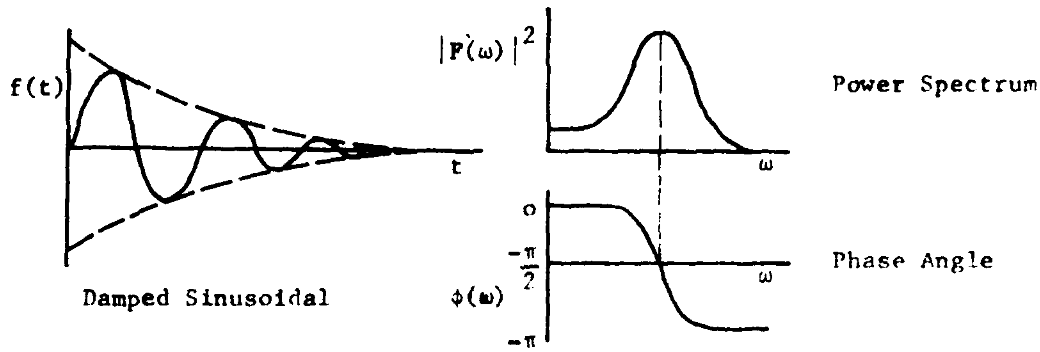


Figure 6. Example of a Transform Pair

The inverse Fourier transform is defined by

$$f(t) = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)e^{j\ \ t}\,d\omega$$

$$Re[f(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\cos\omega t\,d\omega$$

$$Im[f(t)] = \frac{1}{2\pi}\int_{-\infty}^{\infty} F(\omega)\sin\omega t\,d\omega$$

The similarity between the Fourier Transform and the Fourier Series should be noted. It can be shown that discrete spectrum resulting from Fourier Series analysis has the Fourier Transform continuous spectrum as its envelope. See Figure 7.

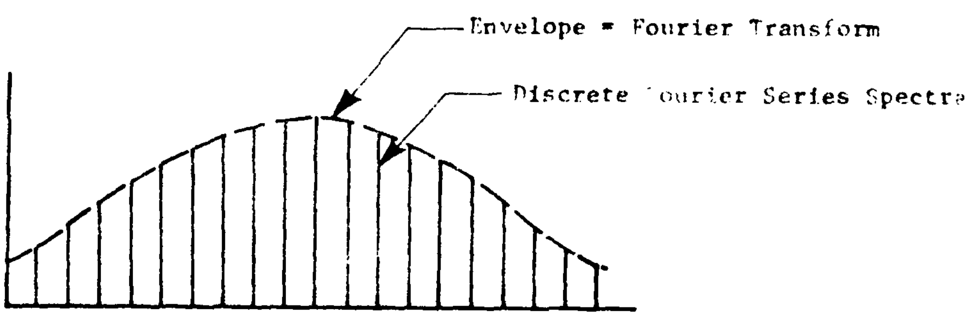


Figure 7. Illustration of the Relationship Between Fourier Series and Fourier Transforms

III. SIGNAL PROCESSING

A. <u>ANALOG TO DIGITAL CONVERSION</u>. There are basically two types of signals encountered in ultrasonics. They are generally called "narrow band" and "broadband". These terms refer to the spectral characteristics of a signal. These concepts will be made clear by the use of a simple correlation that exists between the time domain and the frequency domain.

1. The Fourier spectrum of a continuous wave of frequency, f_o, is a spike located at f_o in the frequency domain. This would be the ultimate narrow band signal. The Fourier spectrum of a single spike in the time domain would be a constant extending from zero to infinite frequency. This would be a perfect broadband pulse. See Figure 8, an illustration of these types of signals. An intuitive correlation that might be made is that the longer the signal duration, the narrower the frequency spectrum. These are the kinds of signals that are generally processed during ultrasonic analyses.

2. The important question is what happens when one has only a finite number of sample points from a signal. The points will be digitally processed to obtain a Fourier spectrum. Is this spectrum a reasonable representation of the true frequency content of the continuous signal? Following the digitization process step-by-step will illustrate some of the problems that do occur.

a. A time signal theoretically has a Fourier transform. Let us track the effects that processing this time signal has on the theoretical or true spectrum. First, the signal is sampled at some rate, say T; that is, data obtained every T seconds. Essentially, the time signal is multiplied by a train of delta functions spaced T seconds apart. In the frequency domain, this corresponds with a pulse train separated by 1/T frequency units. See Figures 9a and 9b.

b. Since only a finite number of samples can be processed, the sampled waveform must be truncated. This is again a multiplication, but this time by a rectangularly shaped function. The transform of such a function is a sinc function. See Figures 9c and 9d. This multiplication also translates into a convolution in the frequency domain. See Figure 9e. The sampled and truncated time domain signal now has a distorted periodic frequency profile. This continuous spectrum must also be sampled and truncated. These results are shown in Figures 9f and 9g.

c. Particular attention should be paid to Figures 9c and 9d. Figure 9 shows that if the sampling rate is not high enough, considerable spectral overlap may occur, thus distorting the true spectrum. The term applied to this suboptimal sampling is called Aliasing. There is a theorem, called the Nyquist sampling theorem, which states if the sampling rate is at least twice the highest frequency contained in the signal, then Aliasing will not occur. See Figure 10e.

d. Figure 9d shows that a rectangular window function has a spectrum with many side lobes in it. Convolving this with periodic spectrum of Figure 9e introduces ripples. The type of distortion is called leakage. This problem has been under study for many years and certain window functions developed that have minimum side lobe energy.

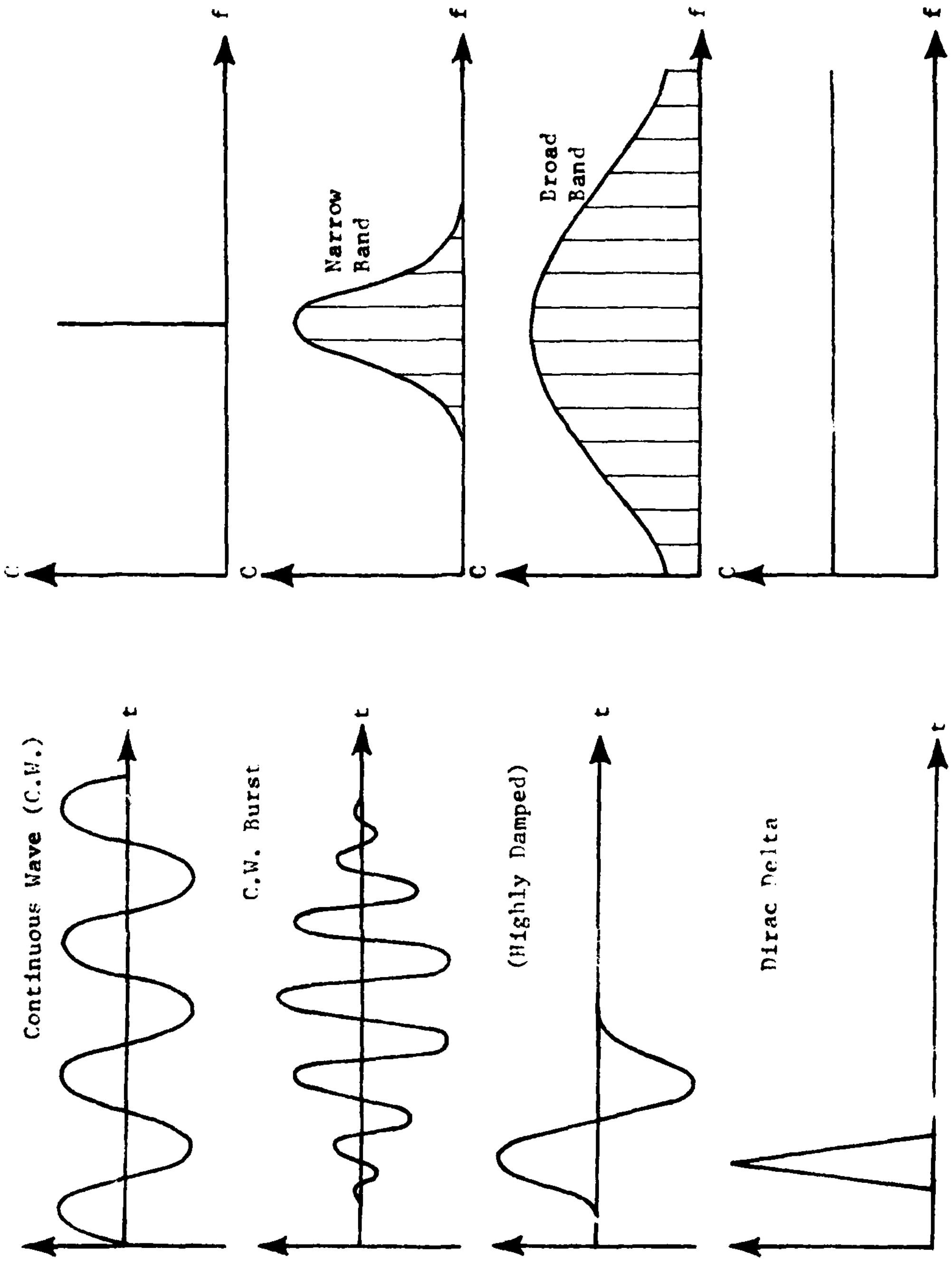


Figure 8. Pulse Types and Corresponding Spectra

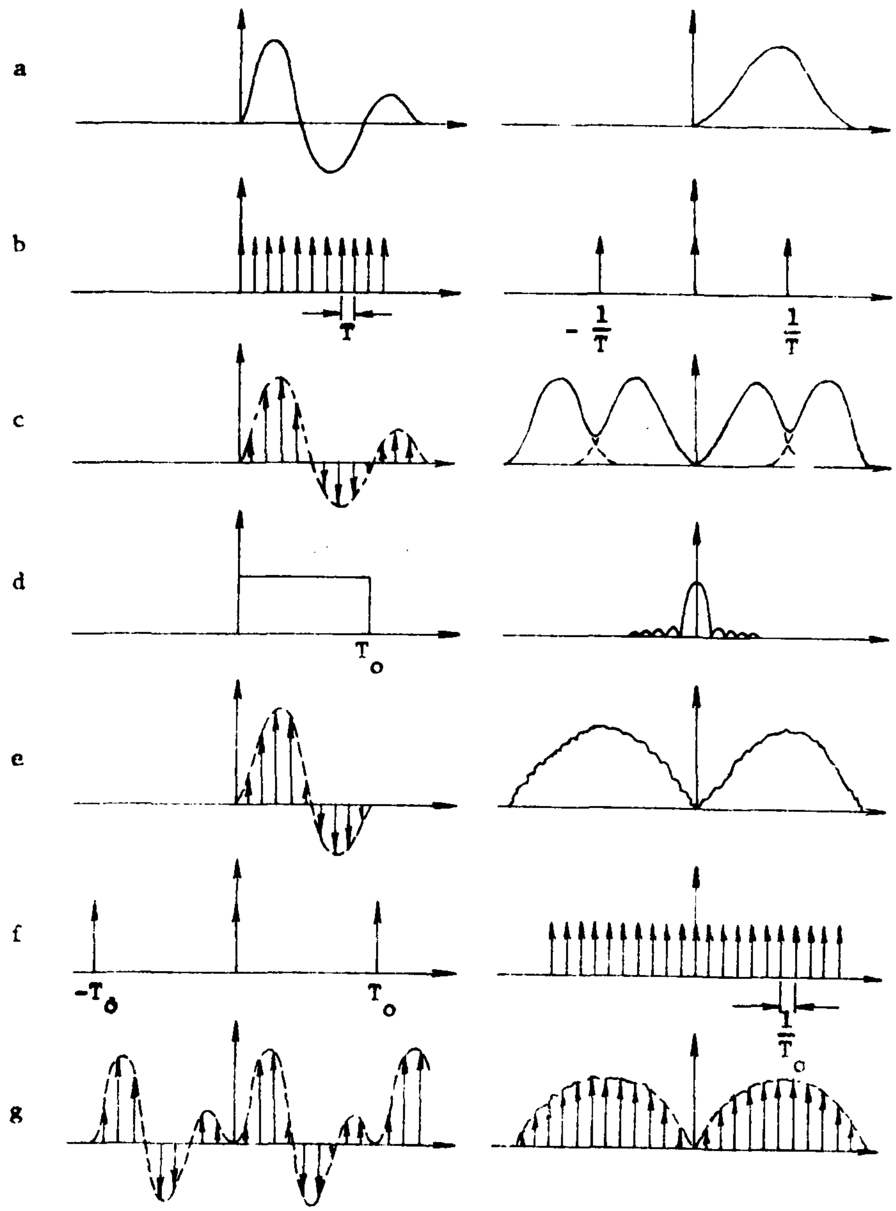


Figure 9. Sampling at an Insufficient Rate

a
b
c
d
e

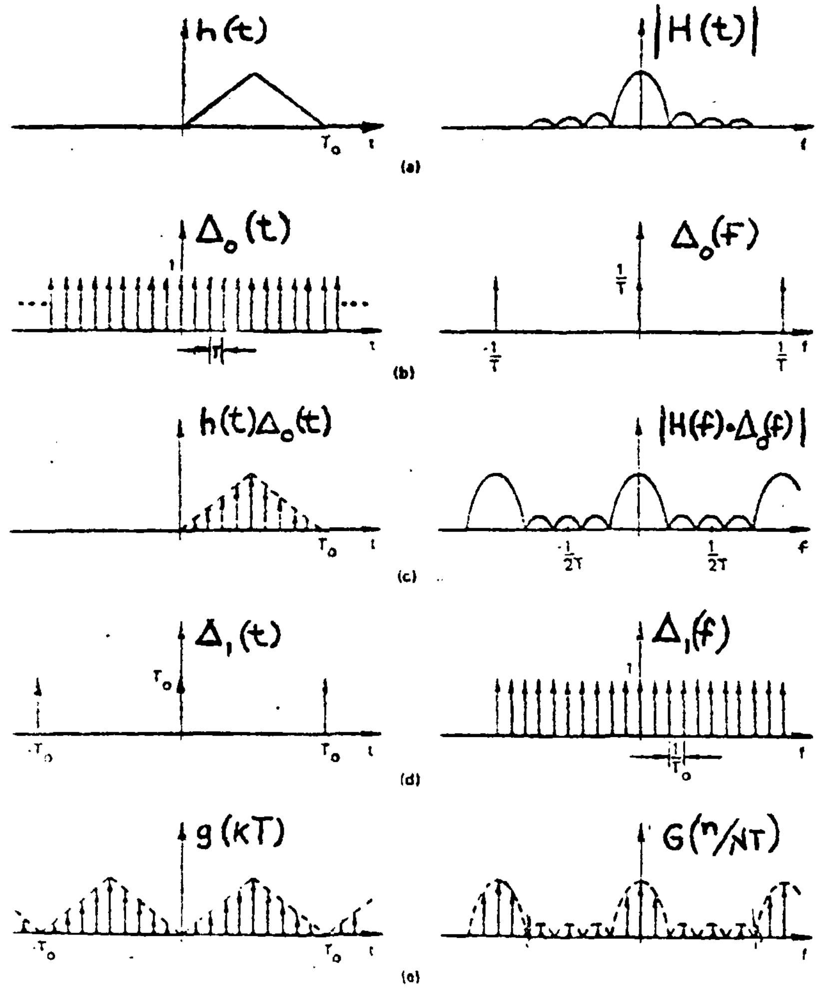


Figure 10. Sampling at a Sufficient Rate

3. Real world signals also have the additional component of noise. Noise may be grouped into several categories. The two items just alluded to, Aliasing and Leakage, might be considered sampling rate noise and mathematical noise. Appropriate steps may be taken to eliminate this type of noise. Quantitization, that is, partitioning the signal into discrete levels, also introduces noise. An analog signal having values located midway between two adjacent quantum levels has a 50-50 chance of being quantitized into either level. Other kinds of noise include electronic noise and thermal noise.

4. Ultrasonic signals may be considered as belonging to the realm of random processes. That is, each time a signal from the same reflector is viewed, it obtains slightly different values. The distribution of these variations may be considered random.

a. One way to get a better estimate of the true value of a signal is to average over a set of similar signals. It can be shown that averaging decreases the effects of quantitization, electronic and thermal noise.

The theory behind simple signal averaging is as follows:

Let x_1 denote an observed signal.

Let s denote the true signal.

Let n_i denote the noise content of the ith observation, then

$$x_i = s + n_i$$

Consider collecting an ensemble of N of these signals:

$$x_1 = s + n_1$$

$$x_2 = s + n_2$$

$$\vdots$$

$$x_N = s + n_N$$

Averaging then involves summing and division by N (simply scaling)

$$\sum_{i=1}^{N} x_i = Ns + \sum_{i=1}^{N} n_i$$

$$\frac{\sum_{i=1}^{N} x_i}{N} = s + \frac{\sum_{i=1}^{N} n_i}{N}$$

For N large enough

$$\sum_{i=1}^{N} n_i$$

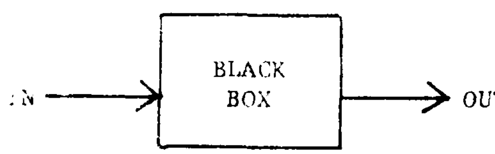
is small, and division by N makes the second right-hand term even smaller:

$$\frac{\sum_{i=1}^{N} x_i}{N} \approx s \text{ for proper } N.$$

b. Averaging introduces the new problem of jitter. Jitter is time shifting of signal components due to the variability of the trigger levels necessary to initiate the digital-to-analog sampling process. Most often, this occurs when instruments are first turned on. After a period of time though, trigger levels tend to stabilize. If jitter is still present, a process called correlation detection may be used.

c. A signal is captured and stored. A new signal is then obtained and cross-correlated with the stored one. The maximum value of the cross-correlation function locates on the time axis the number of sample units the new form has been shifted away from the original. The second is time shifted into agreement with the first one. The two signals are then averaged and stored as a new reference signal. The process may be iterated until it is thought that sufficient noise reduction has occurred.

B. SIGNAL PROCESSING DEFINITIONS. Many textbooks on signal processing are available, many of which could be useful in understanding the difficulties and possible improvements of ultrasonic signal analysis. Highlights and definitions of several terms that are encountered in the signal processing field are outlined below. A review of the terms and basic concepts will serve to introduce the subject and its many mathematical and electrical engineering areas of study. The concept of a transfer function is illustrated in Figure 11. This idea is used very often in signal processing systems and is referred to in the electrical engineering and systems analysis literature. The transfer function is often treated as a black box where the output function can be formulated as a function of an input function by way of the black box or transfer function.



Out = F(In) where black box (F) represents a transfer function.

Figure 11. Transfer Function Concept

1. Signal. A physical disturbance that contains information. The disturbance may vary with time, temperature, pressure, etc. A traffic light is a signal. The color is the disturbance and the information is stop, go, or caution. In ultrasonics, voltage variations versus time are typical signals.

2. Processing. The automatic extraction of information from a signal. For instance, the Fast Fourier Transform is a computer processing technique used to extract the frequency information contained in a signal.

3. Analog Signal. A signal that is continuous in time.

4. Digital Signal. A signal that occurs in discrete time intervals, usually represented as a sequence of numbers, each being restricted to an integer multiple of a fundamental unit called a quantum.

5. Sampling. When it becomes impractical to process a signal in continuous time, samples of the signal are taken at a set of predetermined discrete times.

6. Quantitization. Restriction of sample values to a finite number of possible values.

7. A/D Conversion (Analog to Digital). The procedure for sampling analog signals and thereby converting the analog information into a digital sequence.

8. Spectral Analysis. The evaluation of the frequency content of a signal. This is usually performed by Fourier transforming the signal and noting those areas which have significant values.

9. Bandwidth.

a. The highest frequency above which there is no significant content. (0 to 10 MHz, 0 to 20 MHz, etc.)

b. When prefixed with 3 dB or 6 dB, the width of the spectral profile, at respective amplitudes, 0.707 of the peak value and 0.5 of the peak value respectively. (5 to 10 MHz, 2 to 6 MHz, etc.)

10. Sampling Theorem. A theorem that states an analog signal must be sampled at a rate of $1/(2 f_{max})$, where f_{max} is the highest frequency contained in the signal, in order to insure a faithful representation of the signal in the digital domain.

11. Aliasing. The misinterpretation of a signal due to too low a sampling rate. When one looks at an airplane propeller, it seems to be going slow (or even backwards). This is due to the fact that the eye cannot sample the visual information at a high enough rate. This can produce incorrect electronic signals since a lower sampling rate could actually represent several higher frequency signals. See Figure 12. This can occur when the sampling rate is too low.

12. Filter. A mathematical algorithm or computational procedure used to process digital data. These algorithms are implemented either in software (using computer language) or in hardware (actual digital circuitry).

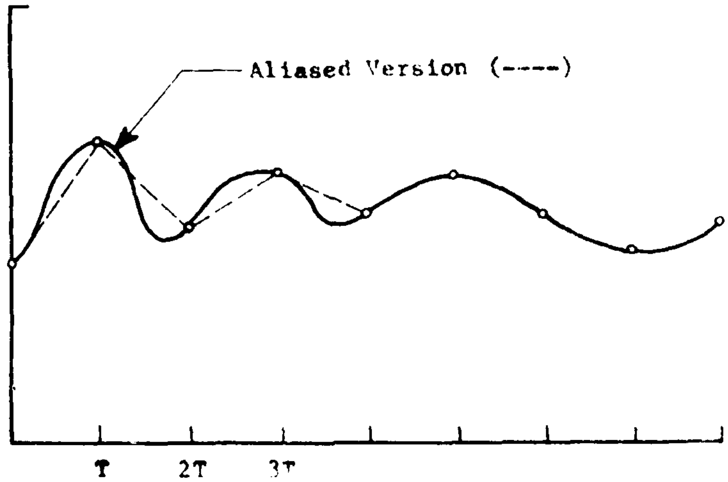


Figure 12. Aliasing

13. Transfer Function. A mathematical representation of the effects a physical system will have on an input, regardless of what that input may be. It represents the inherent characteristics of the system.

14. Linear Filter. A filter where the property of superposition is known to hold. That is, if two inputs are added together, the output is the sum of the two individual outputs obtained from each input alone.

15. Auto Correlation. A statistical measure of the expected value of the product $f(n) \cdot f(n + k)$, where $f(n)$ is a signal at time n, and $f(n + k)$ is the value of the signal k units later. k is called the lag.

$$AC(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)f(t + \tau)dt$$

This function might be useful for the alignment of similar signals displaced relative to each other in time.

16. Power Spectrum. The magnitude of the Fourier spectrum of a signal.

17. Power Spectral Density. The Fourier transform of the auto-correlation of a signal.

18. Signal to Noise Ratio. (A/σ_n) The ratio of peak amplitude to the root mean square of the noise in a signal.

19. White Noise. (Wideband Gaussian Noise) A signal whose power spectrum is a constant.

20. Cross Correlation. A statistical measure of the expected value of the product $f(n) \cdot g(n + k)$ where $f(n)$ is a signal at time n, and $g(n + k)$ is another signal at time n + k.

$$CR(\tau) = \lim_{T \to \infty} \frac{1}{2T} \int_{-T}^{T} f(t)g(t + \tau)dt$$

a. Cross correlation may be useful for the detection and location of a signal which is embedded in noise, since only those components which are not noise will have non-zero components.

b. Intuitively, correlation is similar to a template matching procedure. One signal is displaced relative to another (or the same) and the two compared. That displacement where the two signals agree the most is where the correlation function is maximum and where they coincide the least, it is a minimum.

21. Deconvolution. The process of solving for the function

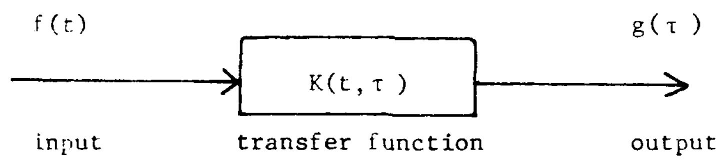
$K(t, \tau)$ in the equation

$$\int_{-\infty}^{\infty} f(t)\, K(t,\tau)\, dt = g(\tau)$$

given the functions $f(t)$ and $g(\tau)$.

$K(t,\tau)$ is known as the kernel of the integral.

a. This is related to the theory of linear systems where it is shown that if $f(t)$ is an input to a system, $g(\tau)$ is the output with $K(t,\tau)$ being the system transfer function.



b. Fourier analysis also shows that if $F(w)$, $K(w)$, and $G(w)$ are the Fourier transforms of $f(t)$, $K(t,\tau)$, and $g(\tau)$ respectively, then

$$K(w) = \frac{G(w)}{F(w)}.$$

This complex division is known as deconvolution. Deconvolution could be useful in ultrasonic analysis, for example, in transducer compensation analysis, that is making one transducer appear to be another, perhaps more suitable transducer.

22. Signal Averaging. A mathematical process (or filter) which extracts the central tendency of a signal. It is usually used to eliminate noise.

IV. BASIC PATTERN RECOGNITION AND CLASSIFICATION PROCESSES

A. SIMPLE TECHNIQUES. One manner in which man has been expanding the general capabilities of the digital computer is the concept of artificial intelligence. Hopefully, the digital computer will perform perceptual tasks assigned to it. Pattern recognition has received considerable attention in this area.

1. The basis for using pattern recognition in solving practical problems lies in the assumption that a logical means exists to train a computer to associate given data with a particular test response.

2. The basic form for a classification process, as illustrated in Figure 13, consists of a data acquisition process, parameter or feature extractor, and a classifier. Note: Before the system functions correctly, the classifier must be trained to provide a solution having a higher probability of being correct than the system previously used.

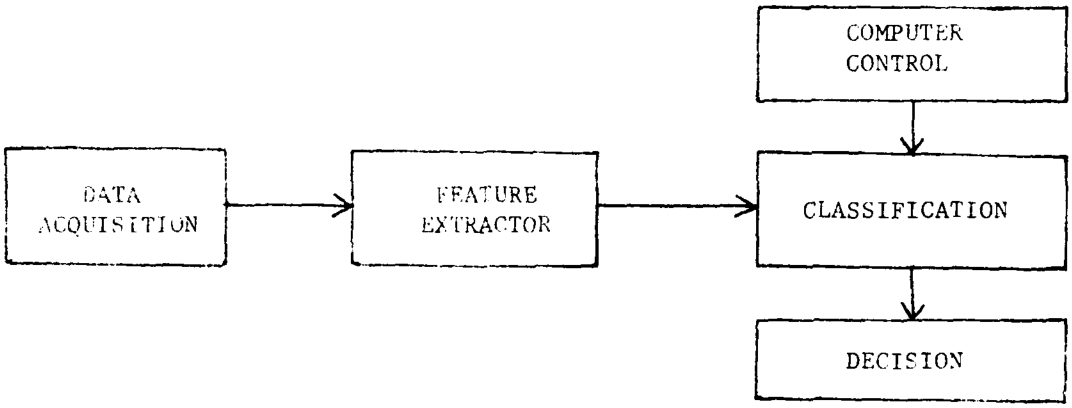


Figure 13. The Pattern Recognition and Classification Process

3. Figure 14 illustrates how Probability Density Function (PDF) curves might be used to obtain feature effectiveness. Note that feature 1 has approximately the same values for the two different classes in the first illustration.

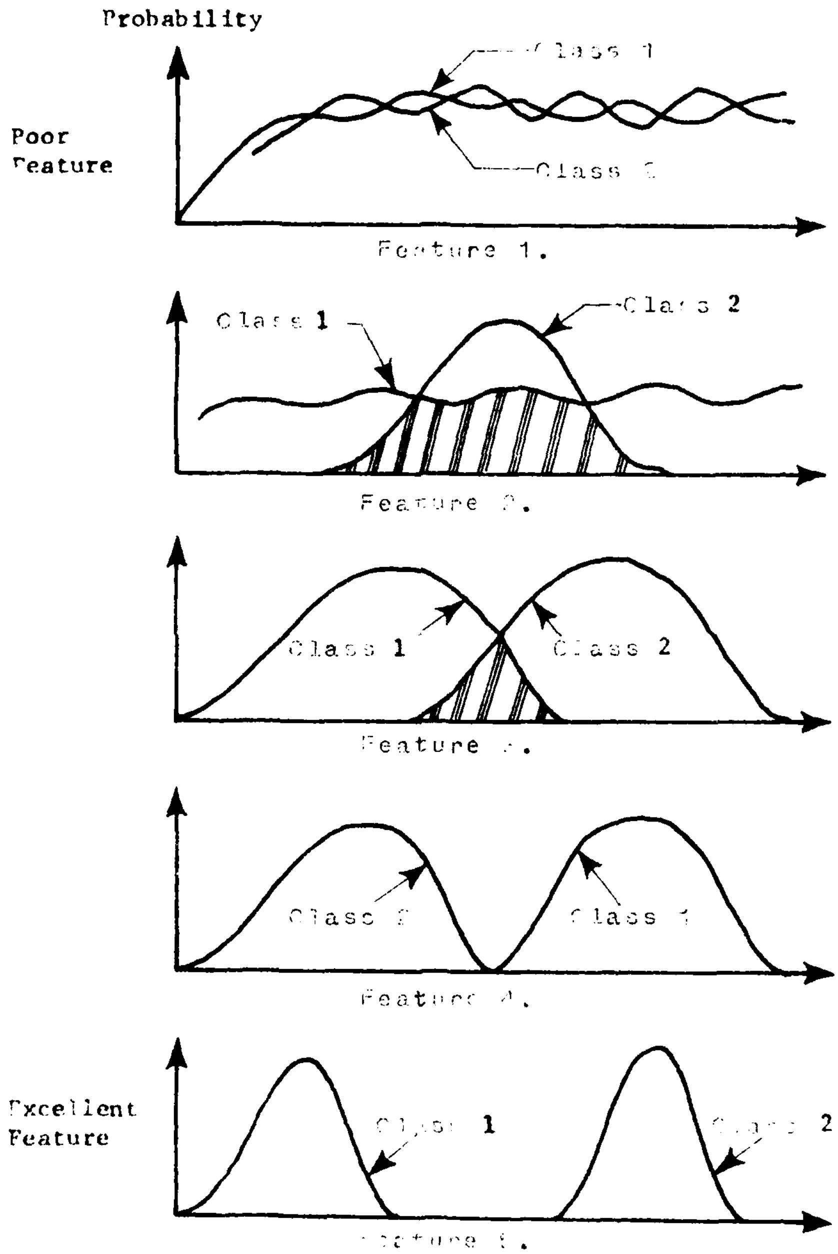


Figure 14. Sample Probability Density Function Curves for a Typical 2 Class Classification Problem

There is no differentiation capability for the feature. Features 2 and 3 are also somewhat limited in their capability for classifying the problem as either class 1 or 2. Note that, however, feature 4, possibly center frequency of the reflected signal provides for us a clearer differentiation between class 1 and class 2, and of course, feature 5, possibly a 6 dB down frequency bandwidth provides for us an excellent feature for differentiating the two classes with 100% reliability. Quite often, however, the features fall in categories such as those illustrated in features 2 and 3. The probability density function curves, however, provide us with insight into the difficulties that might be associated with the classification problem in pattern recognition. Obviously, if results similar to those for feature 5 occur, the solution to the problem is complete. If this is not the case, it is often desirable to examine two-dimensional feature profiles as illustrated in Figure 15.

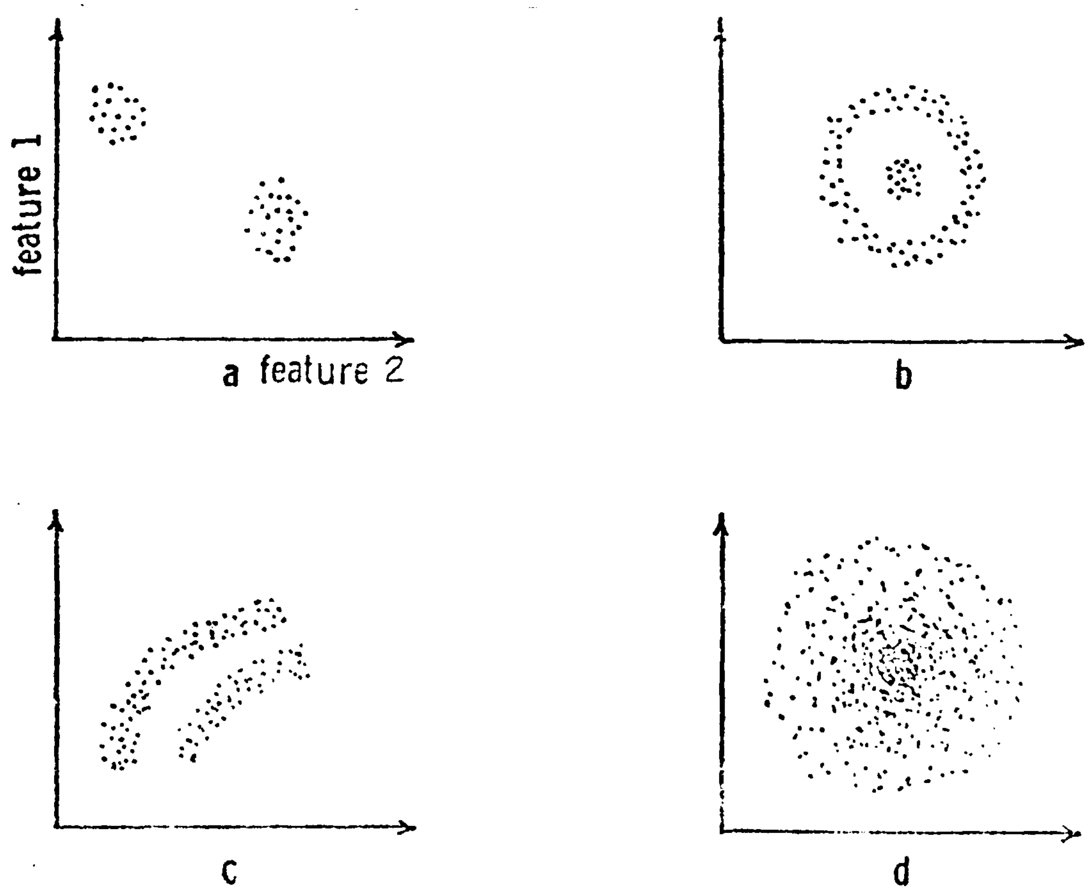


Figure 15. Two-Dimensional Feature Space Clusters

4. If we were to plot in two-dimensional space a feature 1 versus feature 2, and were lucky enough to obtain data clustering as either ball-like, ring type, string type, and so on, as illustrated in Figure 15, again promise for obtaining a reasonable solution exists. The combination of two-dimensional profiles should be plotted for all promising type features as indicated by the probability density function analysis. If the cluster situation in two-dimensional feature space is not useful, it then becomes necessary to employ more sophisticated algorithm analysis from pattern recognition.

5. The next approach that could be used for finding a solution to this classification problem would be to consider aspects of Bayes' decision theory in combination with the results obtained from the probability density function analysis. A fuzzy logic decision algorithm could be established that classifies a certain percentage of the total number of test situations encountered with 100% reliability vector or index of performance. A sample fuzzy logic algorithm is illustrated in Figure 16.

6. If problems are encountered in this approach or a different kind of reliability parameter is required, additional concepts in pattern recognition must be explored. As an example, an index of performance vector that provides us with 100% classification even though the algorithm reliability is only 80% or 90%, could be useful for many applications. Keep in mind that the index of performance criteria depends on the classification levels and possible loss function analysis, loss functions can be incorporated into the index of performance evaluation. As an example, an item classified as class 1 that is really class 2, may not represent a serious error. On the other hand, calling a class 2 situation class 1, could be serious. Suppose we are doing a flaw detection in metals. If class 1 represents porosity and class 2 cracking classification of porosity, cracking is obviously not serious since it results possibly in small financial loss.

5. FISHER LINEAR DISCRIMINANT. One of the major problems encountered in pattern recognition work is the vastness of the feature space. Procedures that are analytically and computationally manageable in low dimensional spaces becomes impractical in higher dimension spaces. An ideal space is the one-dimensional space represented by a straight line. The advantage of a Fisher Linear Discriminant is that it projects all of the data from an N-dimensional space onto the best line for separating the data. Once the data has been projected onto the line, a threshold value may be selected which will separate the data into two classes. Thus the Fisher Linear Discriminant is ideally suited to a two-class problem, as illustrated in Figures 17 and 18.

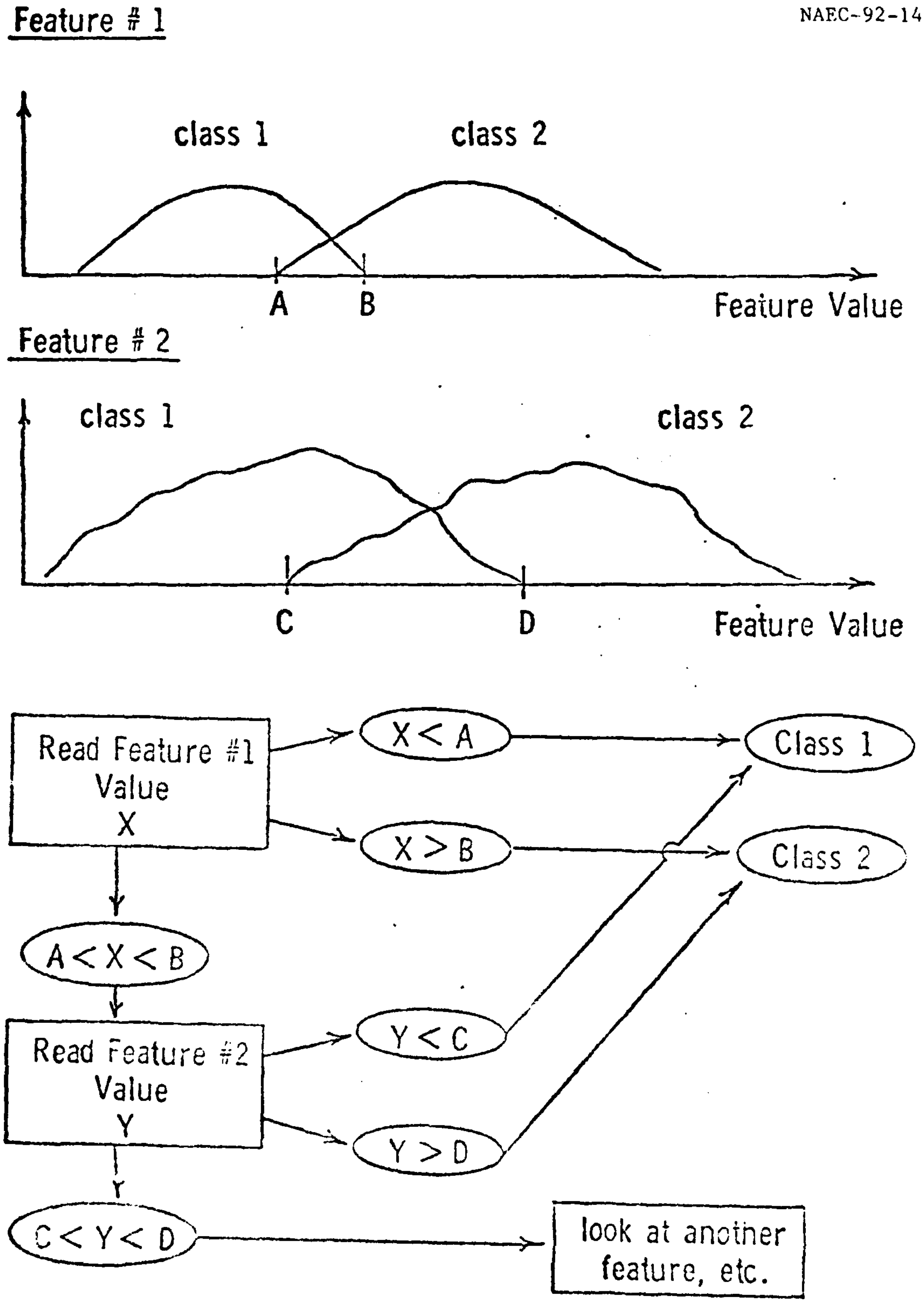
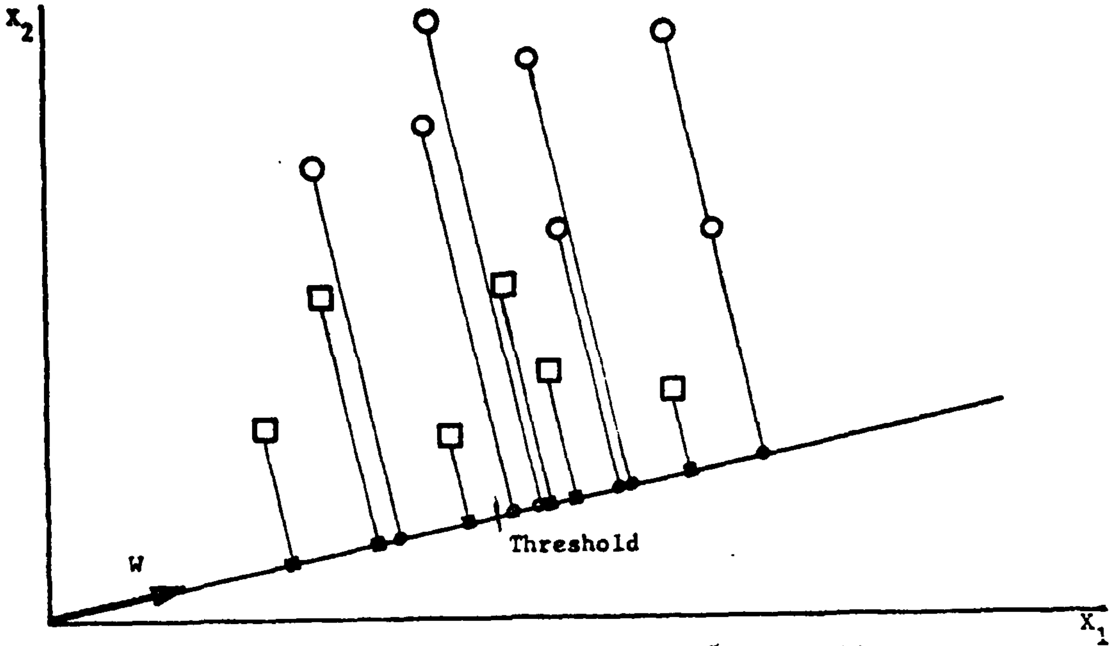
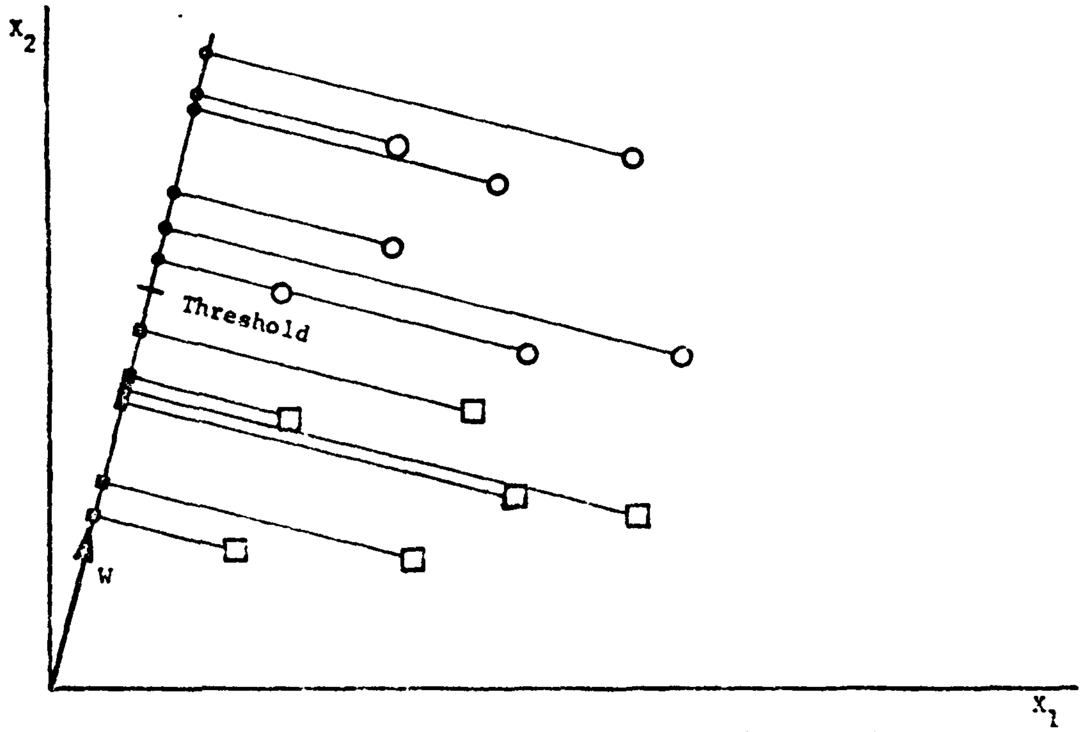


Figure 16. Sample Fuzzy Logic Algorithm Development

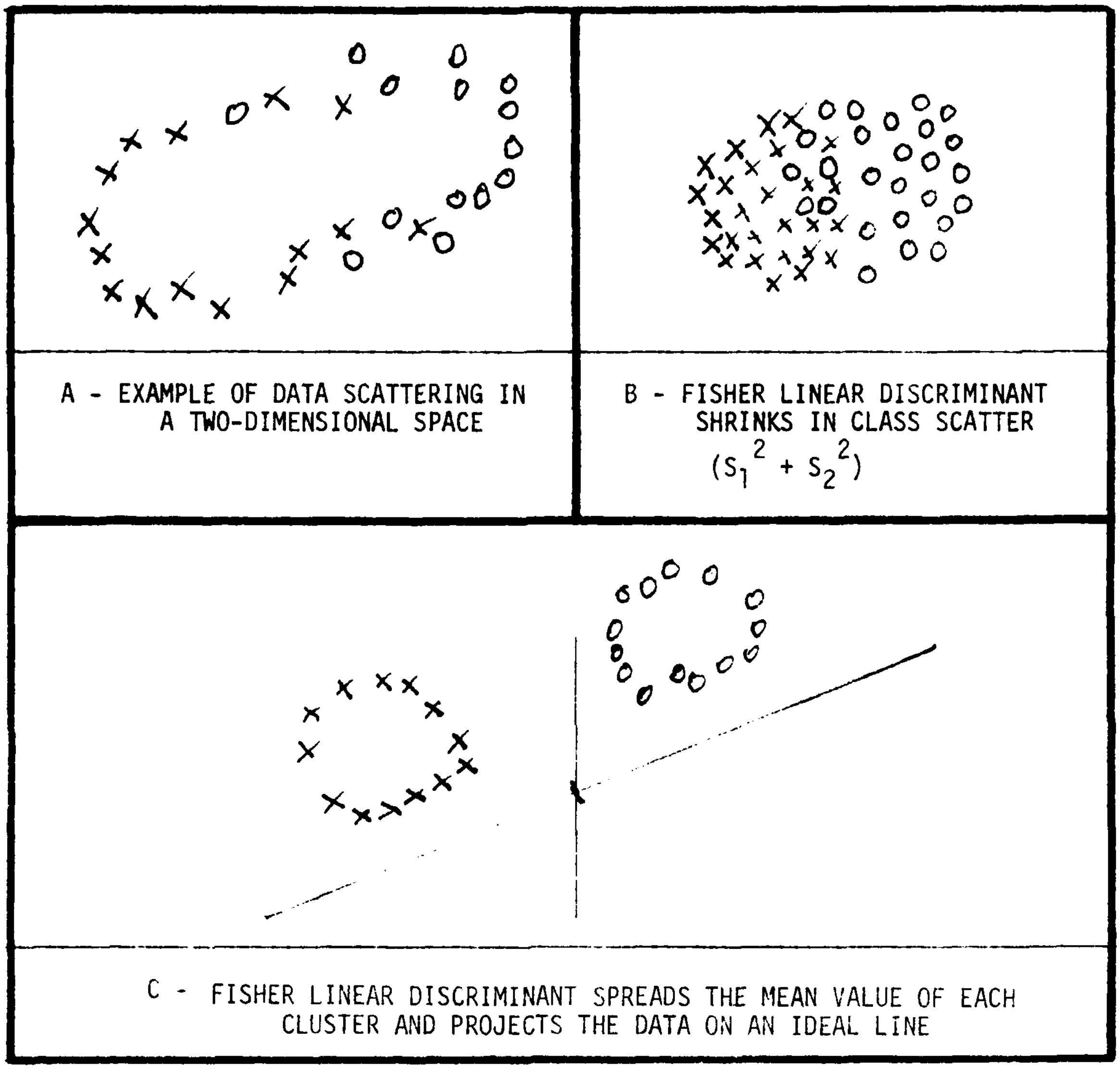


A. Features Projected Onto Arbitrary Line - 50% Reliability



B.

Figure 17. Features Projected Onto Line Determined by the Fisher Linear Discriminant - 100% Reliability



A - EXAMPLE OF DATA SCATTERING IN A TWO-DIMENSIONAL SPACE

B - FISHER LINEAR DISCRIMINANT SHRINKS IN CLASS SCATTER $(S_1^2 + S_2^2)$

C - FISHER LINEAR DISCRIMINANT SPREADS THE MEAN VALUE OF EACH CLUSTER AND PROJECTS THE DATA ON AN IDEAL LINE

Figure 18. Comparison of Two-Dimensional Space and Fisher Linear Discriminant Data Scatter

The simplest way to project an N-dimensional space onto a line is by forming a dot product.

$$\begin{bmatrix}\omega 1, & \omega 2, & \cdots, & \omega_N\end{bmatrix} \cdot \begin{bmatrix}x_1\\x_2\\ \vdots\\ x_N\end{bmatrix} = \sum_{i}^{N} \omega_i x_i = y \text{ (a scalar)}$$

This may be written as

$$y = \underline{w}^t \underline{x}$$

1. Consider a set of K samples (vectors) divided into two classes, C_1 and C_2 with N_1 samples and N_2 samples respectively ($K = N_1 + N_2$). (If the samples fall into two intermingled clusters, the result desired is that the clusters be shrunk and their means well separated.) See Figure 15. Another way of expressing this is to say that the difference of projected means is to be maximized and the scatter within each cluster is to be minimized. The mathematical formulation of this problem is given below.

Let $\underline{x}$ denote a typical D-dimensional vector. Then for class 1 samples, the vector mean is

$$\underline{m}_1 = \sum^{N_1} \underline{x}/_{N_1}, \ \underline{x} \in C_1$$

and for class 2

$$\underline{m}_2 = \sum^{N_2} \underline{x}/_{N_2}, \ \underline{x} \in C_2.$$

The projected means would be

$$\tilde{m}_1 = \frac{1}{N_1} \sum^{N_1} y \qquad y \text{ projected from } C_1$$

$$\tilde{m}_2 = \frac{1}{N_2} \sum^{N_2} y \qquad y \text{ projected from } C_2$$

that is

$$\tilde{m}_1 = \frac{1}{N_1} \sum^{N_1} \underline{w}^t \underline{x} = \underline{w}^t \underline{m}_1$$

the squared difference of projected means is then

$$|\tilde{m}_1 - \tilde{m}_2|^2 = |\underline{w}^t \underline{m}_1 - \underline{w}^t \underline{m}_2|^2 = |\underline{w}^t(\underline{m}_1 - \underline{m}_2)|^2$$

$$= \underline{w}^t(\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - \underline{m}_2)^t \underline{w}$$

this may be rewritten as

$$|\tilde{m}_1 - \tilde{m}_2|^2 = w^t M w$$

where

$$M = (\underline{m}_1 - \underline{m}_2)(\underline{m}_1 - m_2)^t$$

The scatter S for each class may be defined as

$$S_1 = \sum (\underline{x} - \underline{m}_1)(\underline{x} - \underline{m}_1)^t, \ \underline{x} \in C_1$$

$$S_2 = \sum (\underline{x} - \underline{m}_2)(\underline{x} - \underline{m}_2)^t, \ \underline{x} \in C_2$$

Also, a variance measure of the entire data set may be defined as

$$V = \frac{1}{K} (\tilde{S}_1^{\,2} + \tilde{S}_2^{\,2})$$

where

$$\tilde{S}_1^{\,2} = \sum (y - \tilde{m}_1)^2 \qquad y \in \text{projected } C_1$$

$$\tilde{S}_2^{\,2} = \sum (y - \tilde{m}_2)^2 \qquad y \in \text{projected } C_2$$

or

$$\tilde{S}_1^{\,2} = \sum (\underline{w}^t \underline{x} - \underline{w}^t \underline{m}_1)^2 \qquad \underline{x} \in C_1$$

$$= \sum w^t (\underline{x} - \underline{m}_1)(\underline{x} - m_2)^t s \qquad \underline{x} \in C_1$$

$$\tilde{S}_1^{\,2} = \underline{w}^t S_1 \underline{w} \qquad \text{(likewise for } \tilde{S}_2^{\,2}\text{)}$$

with S_1 as above.

Let $S = S_1 + S_2$

then

$$\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2} = \underline{w}^t S_1 \underline{w} + \underline{w}^t S_2 \underline{w}$$

$$= \underline{w}^t (S_1 + S_2) \underline{w}$$

$$\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2} = \underline{w}^t S \underline{w}$$

The Fisher Linear Discriminant is defined as $\underline{w}^t \underline{x}$ for which

$$J(\underline{w}) = \frac{|\tilde{m}_1 - \tilde{m}_2|^2}{\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}}$$

assumes its maximum value. It should be noted that in a sense when $|\tilde{m}_1 - \tilde{m}_2|^2$ is a maximum and $\tilde{s}_1^{\,2} + \tilde{s}_2^{\,2}$ is a minimum $J(\underline{w})$ is maximum.

In terms of the above formulations

$$J(\underline{w}) = \frac{\underline{w}^t M \underline{w}}{\underline{w}^t S \underline{w}}$$

this is known as a generalized Rayliegh quotient. The solution $\underline{w}$ that maximizes $J(\underline{w})$ is given by

$$\underline{w} = S^{-1}(\underline{m}_1 - \underline{m}_2)$$

Computationally, the vector means $\underline{m}_1$ and $\underline{m}_2$ are first calculated. Then the matrices

$$S_1 = \sum (\underline{x} - \underline{m}_1)(\underline{x} - \underline{m}_1)^t \qquad \underline{x} \in C_1$$

$$S_2 = \sum (\underline{x} - m_2)(\underline{x} - \underline{m}_2)^t \qquad \underline{x} \in C_2$$

are calculated and summed to give S. S is inverted and multiplied times the difference of vector means $(\underline{m}_1 - \underline{m}_2)$.

C. INTRODUCTION TO ADAPTIVE LEARNING NETWORKS. This section will develop the tools and concepts necessary for understanding the motivating philosophies behind a learning network. There are essentially four divisions included in this section. The first three will develop the mathematical machinery required for conceptualization of a learning algorithm. The next and most critical section will cover the analogy between human learning and mathematical (computerized) learning.

1. MINIMUM SQUARED ERROR CONCEPTS. Consider the classic problem of finding a line that approximates a set of data in the sense that the sum of the squared errors is a minimum. One wants to find parameters, say a and b such that the line

$$y = ax + b$$

is a good estimate of the inherent functional relationships of data samples, where the samples are given as

$$\begin{array}{ll} s_1, & (x_1, y_1) \\ s_2, & (x_2, y_2) \\ \vdots & \quad\vdots \\ s_N, & (x_N, y_N) \end{array}$$

Given that a and b exist, the error between actual data and the estimate is

$$\text{error}_i = y_i - ax_i - b$$

for each i. The sum of squared errors is

$$\sum_i^N (\text{error}_i)^2 = \sum_i^N (y_i - ax_i - b)^2$$

This last expression is the one that is to be minimized with respect to the parameters a and b. The condition for a minimum is well known from the calculus. It is that the partial derivatives, with respect to the parameters, of the function equal to zero.

$$\begin{aligned} \frac{\partial}{\partial a} \sum_i^N (\text{error}_i)^2 &= \frac{\partial}{\partial a} \sum_i^N (y_i - ax_i - b)^2 \\ &= \sum_i^N 2(y_i - ax_i - b)(-x_i) \\ &= 0. \end{aligned}$$

The -2 is a constant

$$\therefore \quad \sum_i^N (y_i - ax_i - b)x_i = \sum_i^N x_i y_i - a\sum_i^N x_i^2 - b\sum_i^N x_i = 0$$

which can be written as

$$a\sum_i^N x_i^2 + b\sum_i^N x_i = \sum_i^N x_i y_i$$

Likewise

$$\frac{\partial}{\partial b} \sum_i^N (\text{error}_i)^2 \rightarrow \sum_i^N (y_i - ax_i - b)\ (1) = 0$$

$$\text{or} \quad \sum_i^N y_i - a\sum_i^N x_i - Nb = 0,$$

$$a\sum_i^N x_i + b\ N = \sum_i^N y_i$$

Putting the two above equations in matrix form gives

$$\underbrace{\begin{bmatrix} \sum_i^N x_i^2 & \sum_i^N x_i \\ \sum_i^N x_i & N \end{bmatrix}}_{Y} \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\vec{\omega}} = \underbrace{\begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}}_{\vec{b}}$$

If the matrix Y is nonsingular then the solution is

$$\vec{\omega} = Y^{-1}\vec{b}$$

$$\text{or} \quad \underbrace{\begin{bmatrix} a \\ b \end{bmatrix}}_{\vec{\omega}} = \underbrace{\frac{1}{N\sum_i^N x_i^2 - (\sum_i^N x_i)^2} \begin{bmatrix} \sum_i^N x_i^2 & -\sum^N x_i \\ -\sum_i^N x_i & N \end{bmatrix}}_{Y^{-1}} \underbrace{\begin{bmatrix} \sum_i^N x_i y_i \\ \sum_i^N y_i \end{bmatrix}}_{\vec{b};}$$

$$a = \frac{\sum_i^N x_i^2 \cdot \sum_i^N x_i y_i - \sum_i^N x_i \quad y_i}{N \sum_i^N x_i^2 - (\sum_i^N x_i)^2}$$

$$b = \frac{N \sum y_i - \sum x_i \cdot \sum x_i y_i}{N \sum x_i^2 - (\sum x_i)^2}$$

2. GENERALIZED MATRIX APPROACH; PSEUDOINVERSE. A more general problem may be formulated as follows: We want to find the components of a vector ω such that the accumulated squared errors inherent in the expression

$$X\omega = b$$

add to a minimum. Define e = error vector, then $e = X\omega - b$.

Noting the sum of squared errors is also the length of error vectors, we want to minimize

$$||e||^2 = ||X\omega - b||^2 = \sum_i^N (\vec{\omega}^t \vec{x}_i - b_i)^2 \quad (t \equiv \text{transpose})$$

taking partials (denoted by ∇) we have

$$\nabla ||X\vec{\omega} - \vec{b}||^2 = \sum_i^N 2(\vec{\omega}^t x_i - b_i)x_i = 2X^t(X\vec{\omega} - \vec{b})$$

Setting the partials equal to zero.

$$\nabla ||X\vec{\omega} - \vec{b}||^2 = 2X^t(X\vec{\omega} - \vec{b}) = 0$$

$$\rightarrow \quad X^t X\vec{\omega} = X^t \vec{b}$$

If $X^t X$ is non-singular then

$$\vec{\omega} = (X^t X)^{-1} X^t \vec{b}$$

is the solution to our generalized problem. We can write

$$\vec{\omega} = X^+ \vec{b}$$

where $X^+ = (X^t X)^{-1} X^t$ and is called the "Pseudoinverse" of X.

As an example, we will consider the case where an outcome (result) depends on two other variables. For instance,

$$y = \omega_0 + \omega_1 x_1 + \omega_2 x_2$$

with N data samples

$$
\begin{aligned}
&s_1, \quad (x_1^{(1)},\ x_2^{(1)},\ y_1)\\
&s_2, \quad (x_1^{(2)},\ x_2^{(2)},\ y_2)\\
&\vdots\\
&s_N, \quad (x_1^{(N)},\ x_2^{(N)},\ y_N)
\end{aligned}
$$

(the superscripts indicate the sample number) the generalized matrix approach is

$$X\vec{\omega} = \vec{b}$$

$$
\begin{bmatrix}
1 & x_1^{(1)} & x_2^{(1)} \\
1 & x_1^{(2)} & x_2^{(2)} \\
\vdots & \vdots & \vdots \\
1 & x_1^{(N)} & x_2^{(N)}
\end{bmatrix}
\begin{bmatrix}
\omega_o \\ \omega_1 \\ \omega_2
\end{bmatrix}
=
\begin{bmatrix}
y_1 \\ y_2 \\ \vdots \\ y_N
\end{bmatrix}
$$

$$
X^t =
\begin{bmatrix}
1 & 1 & \cdots & 1 \\
x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(N)} \\
x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(N)}
\end{bmatrix}
$$

$$
X^tX =
\begin{bmatrix}
1 & 1 & \cdots & 1 \\
x_1^{(1)} & x_2^{(2)} & \cdots & x_1^{(N)} \\
x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(N)}
\end{bmatrix}
\begin{bmatrix}
1 & x_1^{(1)} & x_2^{(1)} \\
1 & x_1^{(2)} & x_2^{(2)} \\
\vdots & \vdots & \vdots \\
1 & x_1^{(N)} & x_2^{(N)}
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
N & \sum_i^N x_1^{(i)} & \sum_i^N x_2^{(i)} \\
\sum_i^N x_1^{(i)} & \sum_i^N (x_1^2)^{(i)} & \sum_i^N (x_1x_2)^{(i)} \\
\sum_i^N x_2^{(i)} & \sum_i^N (x_1x_2)^{(i)} & \sum_i^N (x_2^2)^{(i)}
\end{bmatrix}
$$

$$X^t b = \begin{bmatrix} 1 & 1 & \cdots & 1 \\ x_1^{(1)} & x_1^{(2)} & \cdots & x_1^{(N)} \\ x_2^{(1)} & x_2^{(2)} & \cdots & x_2^{(N)} \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}$$

$$= \begin{bmatrix} \sum_i^N y_i \\ \sum_i^N x_1^{(i)} y_i \\ \sum_i^N x_2^{(i)} y_i \end{bmatrix}$$

$$\rightarrow \begin{bmatrix} \omega_o \\ \omega_1 \\ \omega_2 \end{bmatrix} = (X^t X)^{-1} \begin{bmatrix} \sum_i^N y_i \\ \sum_i^N x_1^{(i)} y_i \\ \sum_i^N x_2^{(i)} y_i \end{bmatrix}.$$

One should try this approach to the linear curve $y = ax + b$ to gain appreciation for this method.

3. MULTINOMIALS. Linear approximations always remain linear upon composition. As an example, consider the linear device below.

$$\begin{array}{l} x_1 \rightarrow \\ x_2 \rightarrow \end{array} \boxed{\omega_o + \omega_1 x_1 + \omega_2 x_2} \rightarrow y$$

The device implements $y = \omega_o + \omega_1 x_1 + \omega_2 x_2$. These devices may also be used in tandom or in layers.

$$\begin{array}{l} x_1 \rightarrow \\ x_2 \rightarrow \end{array} \boxed{\omega_o + \omega_1 x_1 + \omega_2 y_2} \rightarrow y_1 \rightarrow$$

$$\begin{array}{l} x_1 \rightarrow \\ x_3 \rightarrow \end{array} \boxed{z_o + z_1 x_1 + z_3 x_3} \rightarrow y_2 \rightarrow$$

$$\boxed{p_o + p_1 y_1 + p_2 y_2} \rightarrow \hat{y}$$

$$\hat{y} = p_o + p_1 (\omega_o + \omega_1 x_1 + \omega_2 x_2) + p_2 (z_o + z_1 x_1 + z_3 x_3)$$

$$\hat{y} = (p_o + p_1 \omega_o + p_2 z_o) + (p_1 \omega_1 + p_2 z_1) x_1 + (p_1 \omega_2) x_2 + (p_2 z_3) x_3$$

$$\hat{y} = A + Bx_1 + Cx_2 + Dx_3$$

A, B, C, and D = constants.

The important point here is that the resulting y is still a first order approximation of the functional relationship inherent in the data. The only result to be obtained, regardless of topological structure is

$$y = \sum_i^N a_i x_i$$

where N is the number of parameters or "features" that y depends on.

Data having a relationship involving cross-products and powers of features would be poorly approximated by this scheme. Therefore, non-linear approximations are now considered. The simple case involving 3 features is shown below.

$$\begin{array}{l} x_1 \rightarrow \\ x_2 \rightarrow \end{array} \boxed{\omega_o + \omega_1 x_1 + \omega_2 x_2 + \omega_{12} x_1 x_2} \rightarrow y_{12}$$

$$\begin{array}{l} x_1 \rightarrow \\ x_3 \rightarrow \end{array} \boxed{z_o + z_1 x_1 + z_3 x_3 + z_{13} x_1 x_3} \rightarrow \hat{y}_{13}$$

$$\begin{matrix} y_{12} \rightarrow \\ y_{13} \rightarrow \end{matrix} \boxed{p_o + p_{12}y_{12} + p_{13}y_{13} + p_{123}y_{12}y_{13}} \rightarrow \hat{y}$$

$$\hat{y} = p_o\ p_{12}(\omega_o + \omega_1 x_1 + \omega_2 x_2 + \omega_{12}x_1x_2) + p_{13}(z_o + z_1x_1 + z_3x_3 + z_{13}x_1x_3)$$
$$p_{123}(\omega_o + \omega_1 x_1 + \omega_2 x_2 + \omega_{12}x_1x_2)(z_o + z_1x_1 + z_3x_3 + z_{13}x_1x_3)$$

which implies terms in

$$x_1,\ x_2,\ x_3,\ x_1x_2,\ x_1x_3$$

as expected and the new interactions

$$x_2x_3,\ x_1x_2x_3,\ x_1^{\ 2},\ x_1^{\ 2}x_2x_3$$

It is noted that cross-terms and power terms are automatically introduced by layering. This suggests that the inclusion of non-linear terms and layering gives a broader range of approximating power.

4. BASIC ALN CONCEPTS. The nonmathematical concept of adaptive learning is easily comprehended. Experiences are recorded by an individual and are put into a scheme or logic by some undefined method. He has a theory about his experiences. When presented with new experiences, his theories are tested. Some theories are modified, some are disregarded, and others remain unchanged. In this sense the individual "adapts" to his environment. The measure of how well his theories perform is the frequency with which he makes successful decisions on new experience.

a. The jump from the philosophical domain to the mathematical one is made most easily by defining terms or creating a vocabulary.

(1) Feature - a quantitative measure of an experience. Examples: temperature, velocity, mean value, peak-to-peak amplitudes, etc.

(2) Feature Vector - a column-like array of features.

$$\begin{bmatrix} \text{temperature} \\ \text{velocity} \\ \cdot \\ \cdot \\ \cdot \\ \text{mean value} \\ \cdot \\ \cdot \\ \cdot \\ \text{frequency} \end{bmatrix}$$

b. Mathematically, once the parameters are defined they may be tagged.

x_1 - temperature

x_2 - velocity

x_{22} - mean value

x_N - frequency

The feature vector may be written as

$$\begin{bmatrix} x_1 \\ x_2 \\ \cdot \\ \cdot \\ x_{22} \\ \cdot \\ \cdot \\ \cdot \\ x_N \end{bmatrix}$$

Other terms will be defined as necessary. From the previous section, we have good reason to suspect that the use of nonlinear expressions and layering will lead to much broader theories than linear relationships. We will restrict ourselves to expressions of the form.

$$y_{ij} = \omega_o + \omega_1 x_i + \omega_2 x_j + \omega_3 x_i x_j + \omega_4 x_i^2 + \omega_5 x_j^2$$

where $i \neq j$. We will be interested in all experiences or features at the outset because we are not aware of any particular relationship.

c. This is best done by considering all possible combinations of features. Given N features, there will be

$$\frac{N(N-1)}{2}$$

possible combinations. As an example, consider $N = 4$. A typical feature vector would be

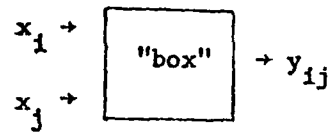
$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \\ x_4 \end{bmatrix}$$

and the combinations would be

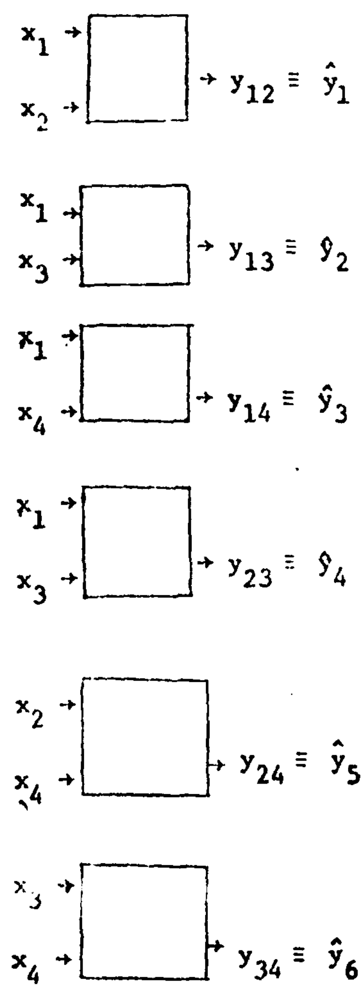
$$x_1\, x_2\,,\ x_2\, x_3\,,\ x_3\, x_4\,,\quad x_1\, x_3\,,\ x_2\, x_4\,,\quad x_1\, x_4 \qquad \text{or } \frac{4(3)}{2} = 6 \text{ combinations}$$

We define a "box" as a device that implements

$$y_{ij} = \omega_o + \omega_1 x_i + \omega_2 x_j + \omega_3 x_i x_j + \omega_4 x_i^{\,2} + \omega_5 x_j^{\,2}$$



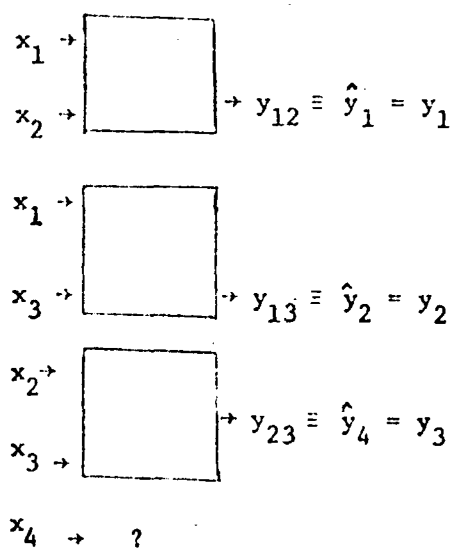
then as a first attempt at developing a theory we have



This initial try will be called a "1st layer". The data that is put into this layer is called "fitting" data. The word fitting referring to the previously developed method that solved for $\omega_0, \omega_1, \ldots, \omega_5$ (pseudoinverse, least squares concepts).

d. The data is submitted to each box in the 1st layer and the coefficients for each box are solved for using the matrix methods at the beginning of this section.

e. The expression or algorithm we have, along with the coefficients, constitute our initial theories about our experiences (features). To decide which theories ("boxes") are useful or not, we must subject them to new experiences. This set of data is called the "selection" set. Those boxes which perform well are retained, while those that perform poorly are disregarded. Assume that boxes 3, 5, and 6 performed poorly. Then we have



We note that feature x_4 may be disregarded at this 1st layer, although the possibility exists that it might be useful at another level in the network.

D. FACTOR ANALYSIS IN PATTERN RECOGNITION. As part of the pattern recognition capability being developed at Drexel University for the Navy, we have incorporated a factor analysis program to efficiently select the features most crucial in the flaw detection problem. This program aids in the identification of relationships among the variables and may contribute to the discovery of new features which will improve our ability to discriminate between different pattern classes. Also, through the use of this program we may be able to reduce the number of measurements needed to make a successful discrimination.

1. FACTOR ANALYSIS. Factor analysis is an extension of principal component analysis which determines the minimum number of independent dimensions needed to account for most of the variance in an original set of variables. Factors are

derived measurement constructions which may produce parsimony[1], orthogonality, increased reliability and increased normality over the observation measures from which they are derived.

a. The digitized ultrasonic waveform can be represented as a signal vector in an n dimensional time space, where each dimension corresponds to the signal voltage at a different latency[2] point in the analysis epoch. Principal component analysis can identify the actual dimensionality of the "signal space" containing a set of such vectors representing waveforms from many derivations in the same experiment or from the same derivations in many experiments. One can then construct a parsimonious description of each waveform as a linear combination of a set of terms. Each term defines the relative contribution of each feature to that waveform. These linear equations enable great data compression, since any waveform in that signal space can be described as some combination of the same basic factors. Thus, patterns of factor weightings can be used to construct clusters of waveforms with distinctive morphology. The p linear combinations of the variables (principal components) are designed to capture as much of the variation in the data as possible while at the same time being linearly independent of all the other principal components.

b. A principal component Y_j is a linear combination of p variables. Thus

$$Y_j = B_1X_{1j} + B_2X_{2j} + \ldots + B_pX_{pj}, \; j = 1, 2, \ldots, m$$

is a principal component with unknown coefficients $B_1, B_2, \ldots, B_p$. In matrix notation let

$$\underline{B} = \begin{bmatrix} B_1 \\ B_2 \\ \cdot \\ \cdot \\ \cdot \\ B_p \end{bmatrix}, \quad \underline{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ \cdot \\ \cdot \\ \cdot \\ Y_m \end{bmatrix}, \text{ and } \underline{X} = \begin{bmatrix} X_{11}, & \ldots, & X_{p1} \\ X_{12}, & \ldots, & X_{p2} \\ \cdot & & \\ \cdot & & \\ \cdot & & \\ X_{1m}, & \ldots, & X_{pm} \end{bmatrix}$$

Then we can write the principal component as

$$\underline{Y} = \underline{XB}$$

1. dimensionality reduction

2. not yet apparent, but there

For a given $\underline{B}$, the sample variance of $\underline{Y}$ is given by

$$\text{var } \underline{Y} = \underline{B}'\underline{SB}$$

where $\underline{S}$ is the sample covariance matrix.

c. The first problem in principal component analysis is to find the principal component, $\underline{Y}_1$, with the maximum variance. The problem then is

$$\text{maximize } \underline{B}'\underline{S}\underline{B} \text{ subject to } \underline{B}'\underline{B} = 1$$

d. If we let

$$\phi = \underline{B}'\underline{SB} - \lambda \, (B'B - 1)$$

where λ is a Lagrange multiplier, the vector of partial derivatives is

$$\frac{\partial\phi}{\partial\underline{B}} = 2\underline{SB} - 2\,\lambda\underline{B}$$

which, upon being set to zero, reduces to

$$(\underline{S} - \lambda\underline{I})\underline{B} = 0$$

To solve this equation, we find the p characteristics roots of the covariance matrix S, thus to maximize the variance of $\underline{Y}$, we choose the largest characteristics root of the covariance matrix S. The first principal component is given by

$$\underline{Y}_1 = \underline{XB}_1$$

with variance equal to λ_1.

e. In general, when there are p variables, the first principal component $\underline{Y}_1$, is a linear combination of the p variables with coefficients equal to the normalized characteristic vector associated with the largest characteristic root of S. The second principal component, $\underline{Y}_2$, is the linear combination of the p variables with coefficients equal to the normalized characteristic vector associated with the second largest characteristic root of S, and so forth up to the pth principal component, Y_p. Each principal component has variance equal to its corresponding characteristic root and each component merely defines the p axes of the p-dimensional concentration ellipsoid and is computed by the program.

f. Thus far in the development of the program, we have the standard packages to accomplish the following:

- CORRE - to find means, standard deviation, and the correlation matrix
- EIGEN - to compute eigenvalues and associated eigenvectors of the correlation matrix
- TRACE - to select the eigenvalues that are greater than or equal to the control value specified by the user
- LOAD - to compute a factor matrix
- VARMX - to perform varimax rotation of the factor matrix

The program has been debugged and employed on a small set of data. The preliminary results are described in the pages which follow.

2. RESULTS AND DISCUSSION. The results of the factor analysis are summarized in the matrix of common factor coefficients presented in Table II. Each entry a_{ij} of the matrix shows the importance of the influence of factor j upon variable i. The factor loadings indicate the net correlation between each factor and the observed variables or features.

a. The interpretation of factor loadings may also be made in terms of the squares of the coefficients. Each $(a_{ij})^2$ represents the proportion of the total unit variance of variable i which is explained by factor j, after allowing for the contributions of the other factors. Thus in the first row of the table, it can be seen that 90% of the variation in Feature 1 can be explained by Factor 1. Factor 2 explains only 0.8%; Factor 3, 2.9%; etc.

b. The matrix of factor loadings, in addition to indicating the weight of each factor in explaining the observed variation, provides the basis for grouping the features into common factors. Each feature may reasonably be assigned to that factor in which it has the highest loading. Where loadings of a feature in two factors are very close, the feature is assigned to the one judged to have the closest affinity. In Table II clusters of features with highest factor loadings are enclosed in rectangles. Only Factor 1 contains a cluster of features.

c. From an examination of the variables in each cluster it appears reasonable to assign attributes to each factor. Thus, Factor 1 might be termed the "Frequency Factor."

d. The analysis described is an exploratory one and, therefore, any conclusions resulting must be tentative. However, the loadings are small in Factor 5; this factor may be eliminated completely by changing the minimum eigenvalue to be retained. In addition, one might conclude from Factor 1 that only one of the frequency features is heavily loaded on the same factor. This conclusion is supported by a previous study conducted by the authors.

e. The generalized variance is shown in Table III for each of the five factors. More than 82% of the generalized variance can be attributed to the first two factors. Thus, it appears that the fractional power ratio at 2-2.5 MHz and the total power from 0-3 MHz are not strong discriminators since their heaviest loadings are on factors other than the first two.

f. The correlation coefficients are shown in Table IV where 10 dB down bandwidth is shown to have a strong negative correlation with 10 dB down mid-frequency, -.928. This correlation is reflected in the cluster of these two features in Factor 1 of the rotated factor matrix in Table II.

g. While the work reported here is preliminary and results must be treated cautiously because of small sample sizes, it appears that factor analysis is a powerful tool for further use in guiding the selection of features from ultrasonic waveforms. This capability should greatly enhance the possibility of selecting features in the most efficient manner for discrimination.

TABLE II. ROTATED FACTOR MATRIX

FEATURE	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
10 dB Down Mid-Frequency	-.949	.091	$\boxed{.170}$	-.189	.163
10dB Down Bandwidth	$\boxed{.941}$	.058	-.249	.142	.172
Number of Peaks Above 20 dB	-.022	-.986	.034	$\boxed{.162}$	-.009
Fractional Power Ratio 2-2.5 MHz	$\boxed{.249}$	.013	-.957	.147	.003
Total Power 0-3 MHz	.408	-.498	-.305	$\boxed{.701}$	-.001

TABLE III. GENERALIZED VARIANCE FOR FACTORS

	FACTOR 1	FACTOR 2	FACTOR 3	FACTOR 4	FACTOR 5
Eigenvalues	2.817	1.295	.677	.157	.055
% of Generalized Variance	56%	26%	14%	3%	1%
Cumulative % of Generalized Variance	56%	82%	96%	99%	100%

TABLE IV. CORRELATION COEFFICIENTS

	NUMBER OF PEAKS ABOVE 20 dB	FRACTIONAL POWER RATIO 2-2.5 MHz	TOTAL POWER 0-3 MHz	10 dB DOWN BANDWIDTH	10 dB DOWN MID-FREQUENCY
Number of Peaks Above 20 dB	1.000	-.027	.586	-.65	-.096
Fractional Power Ratio 2-2.5 MHz	-.027	1.000	.491	.495	-.425
Total Power 0-3 MHz	.586	.491	1.000	.530	-.617
10 dB Down Bandwidth	-.065	.495	.530	1.000	-.928
10 dB Down Mid-Frequency	-.096	-.425	-.617	-.928	1.000

V. GANPUI COMPONENTS

A. DATA ACQUISITION. Shown in Figure 19 is a block diagram of the basic hardware components used by GANPUI. The system generally has a capability of accurately digitizing signals up to the range of 10 MHz. All functions are controllable either manually or through software. Associated with the video terminal, but not shown, is a hard copy unit; that is, a device which reproduces on paper the contents of the video screen. There are also three means of "soft" storage: disk, floppy disk, and cassette.

1. The ultrasonic investigator first determines the appropriate mode of inspection, either contact or immersion. If he chooses immersion, software control for the x-y scanner apparatus is available to him.

2. He then sets such variables as damping, gain, repetition rate, and so forth on the instrumentation involved. Once he is satisfied that proper signals are being obtained, he is ready to use GANPUI.

3. The principal data acquisition program is called GETTER. This program allows the operator to choose the sampling rate, the time window, and number of times a signal is to be averaged, if such a procedure is deemed necessary. The operator may also assign a name to the data he is acquiring. Each set of data is called a frame. Data is automatically plotted on the video terminal as it is obtained. This allows the operator to monitor the procedure and detect any gross jitter (time shifting) problems. The investigator then decides if the data is acceptable for further processing or not, his decision being based on quantum levels involved, noise content, etc. Figure 20 shows a typical format of the GETTER output. Acceptable data is then displayed as an averaged waveform and as a Fourier transformed signal. The information pertinent to the experiment is stored and printed on the second page. There is also space available for comments. Both the averaged waveform and its spectrum are stored. If the operator wishes to continue, another frame is available to him. Upon completing the desired sequence of data, the video terminal displays the names of the files where the data is stored. Figures 21 through 25 show typical outputs.

4. When jitter becomes a significant problem, the operator may employ a correlation-detection algorithm. This procedure aligns signals according to their degree of correlation with each other. After the signals are aligned, they are averaged.

5. The GANPUI system also includes procedures for determining the sensitivity requirements of a particular inspection. Testing sensitivity is determined by the minimum size of a flaw which must be detected according to pertinent specification or other engineering requirements. If the sensitivity is too low, it is possible to miss flaws which are dangerous for the structural strength. Too much sensitivity causes detection of the great amounts of structural inhomogeneities and insignificant flaws.

6. The section on Signal Processing includes descriptions of averaging and correlation detection.

7. The program that retrieves previously acquired data for further processing to evaluation is called REPLAY. The input to this program is the name of the file containing the desired data. The operator may select only those frames which he feels are useful and disregard those that are not of interest. See Figures 26 through 33.

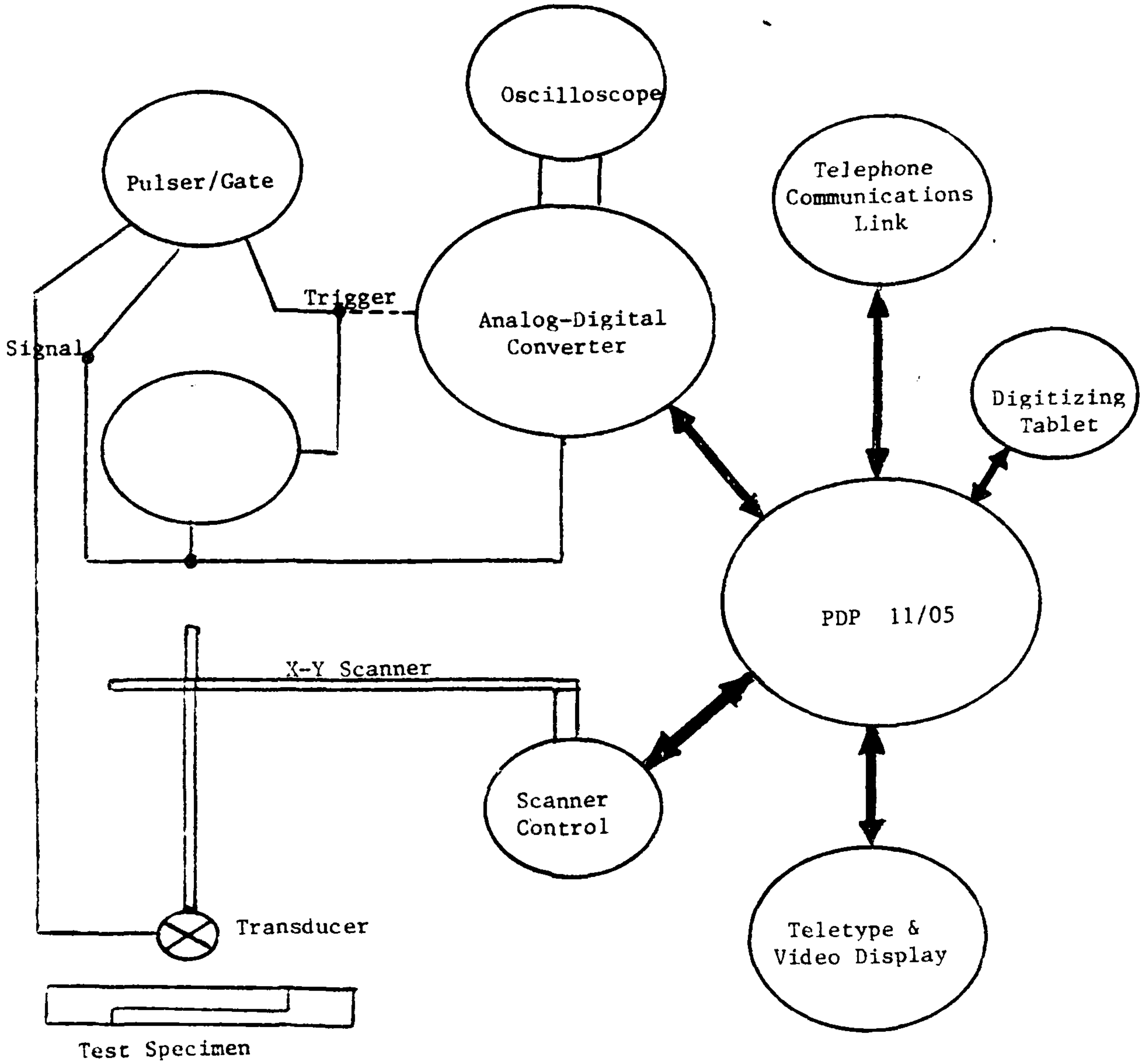


Figure 19. Block Diagram of a Fast Ultrasonic Data Acquisition and Analysis System

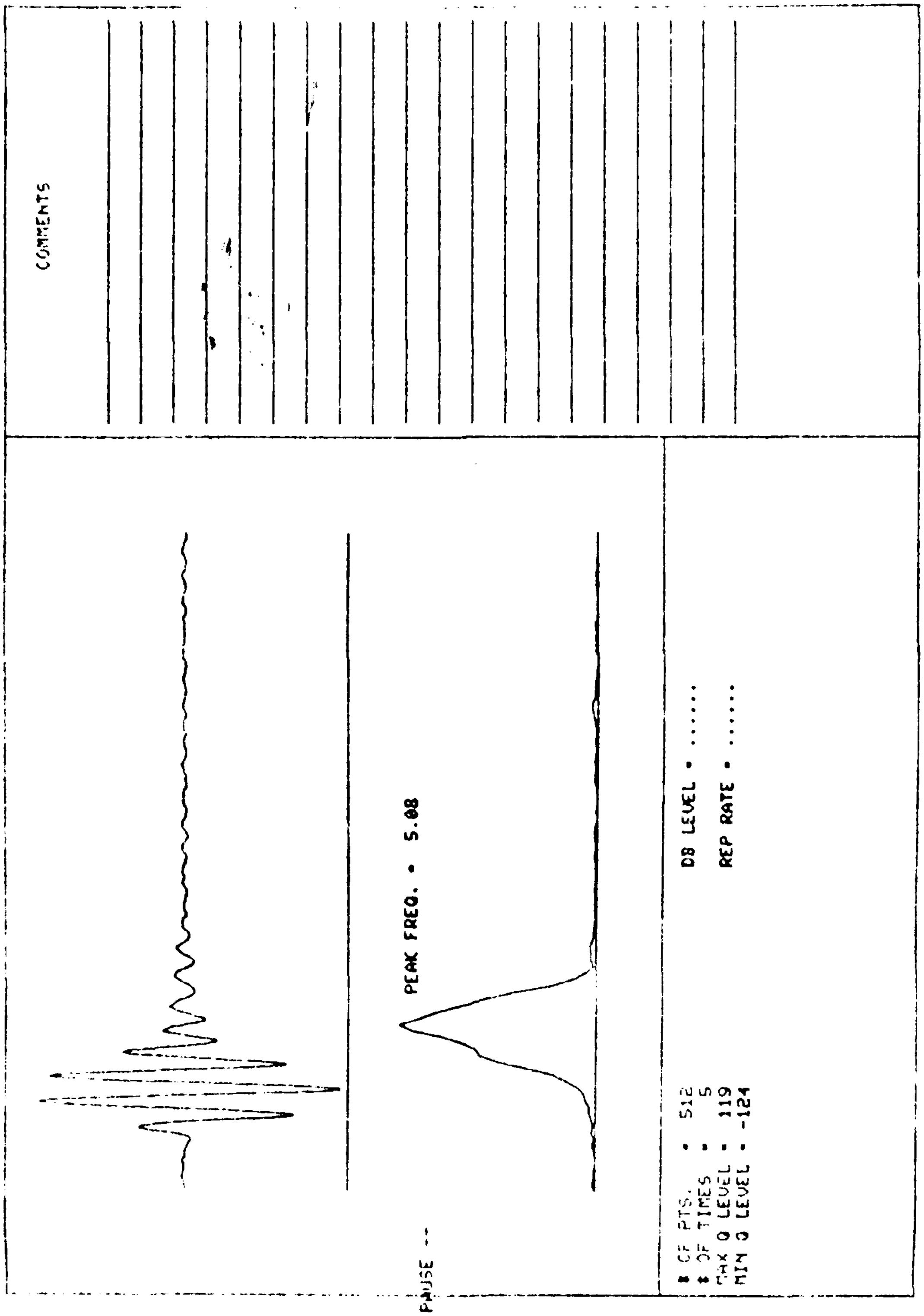


Figure 20. Format of the Getter Output

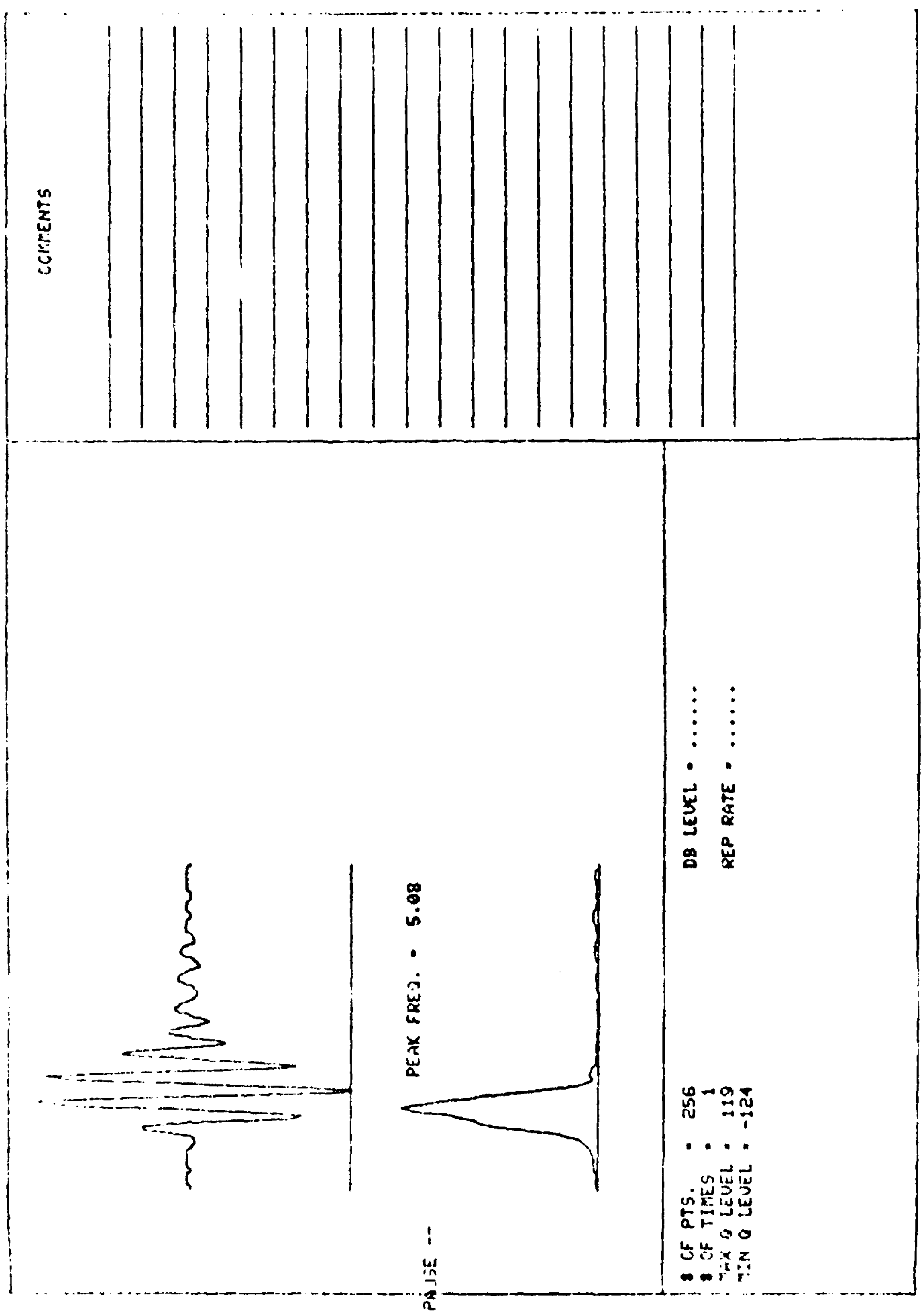


Figure 21. Typical Output Getter Frame

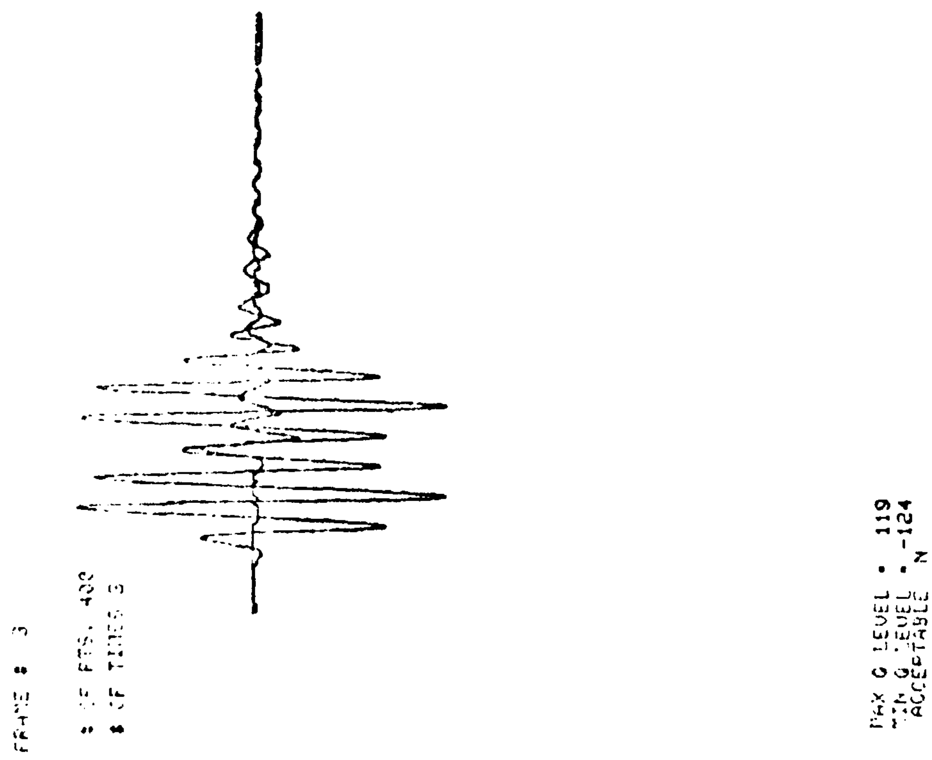


Figure 22. Typical Output Getter Frame

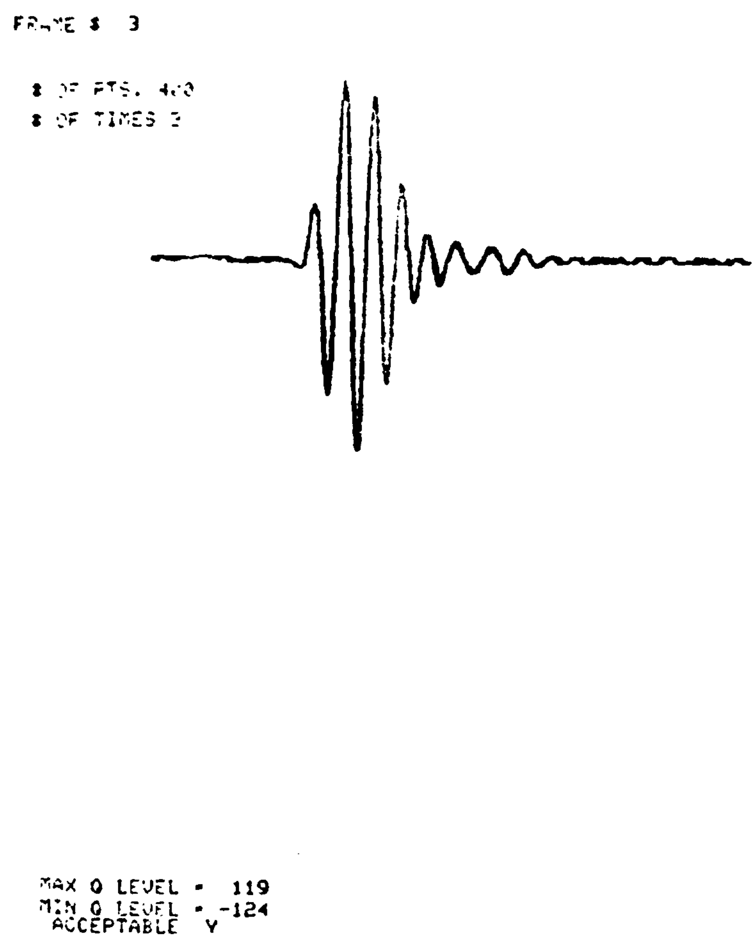


Figure 23. Typical Output Getter Frame

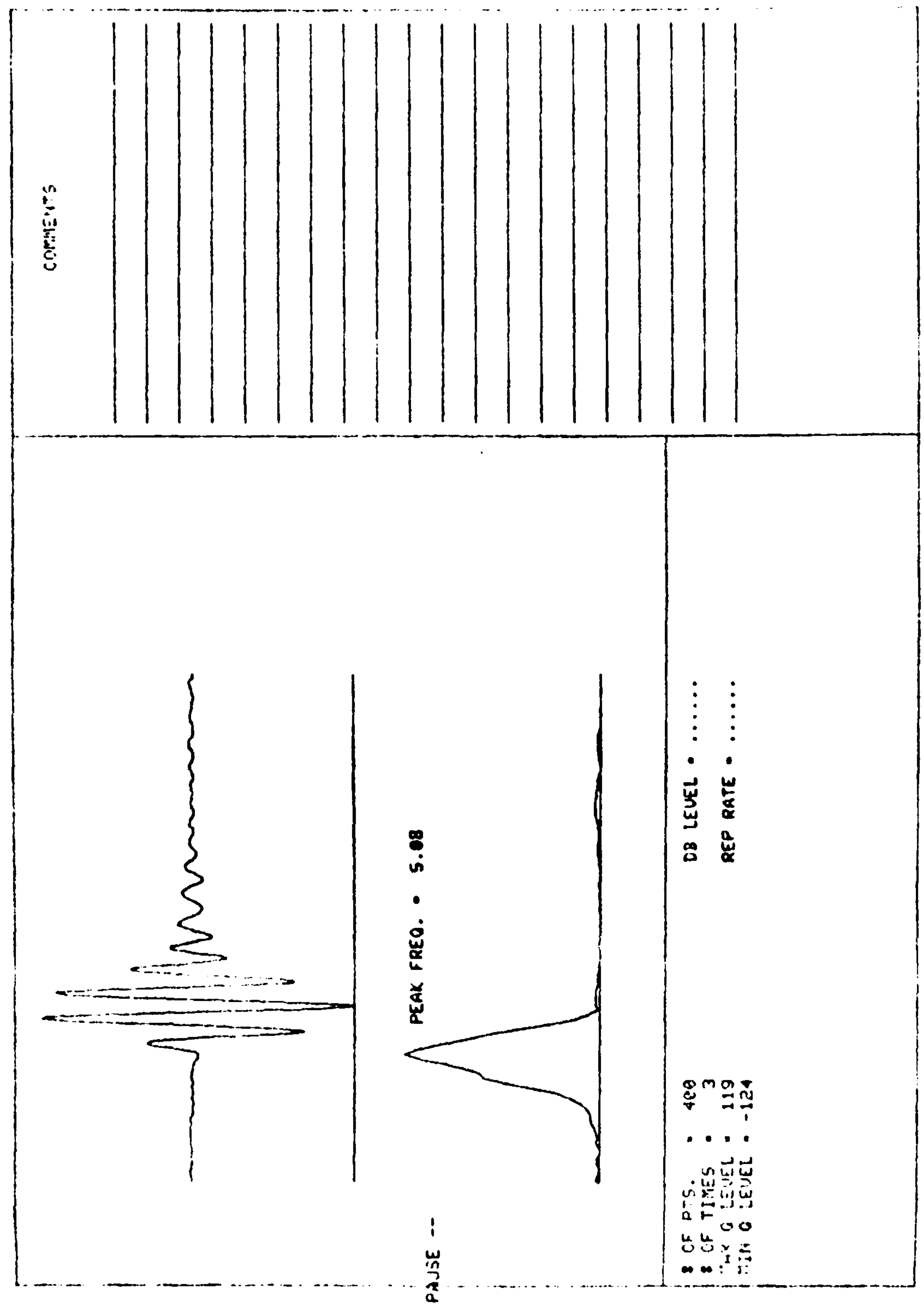


Figure 24. Typical Output Getter Frame

Figure 25. Blank Frame

Figure 26. Blank Frame

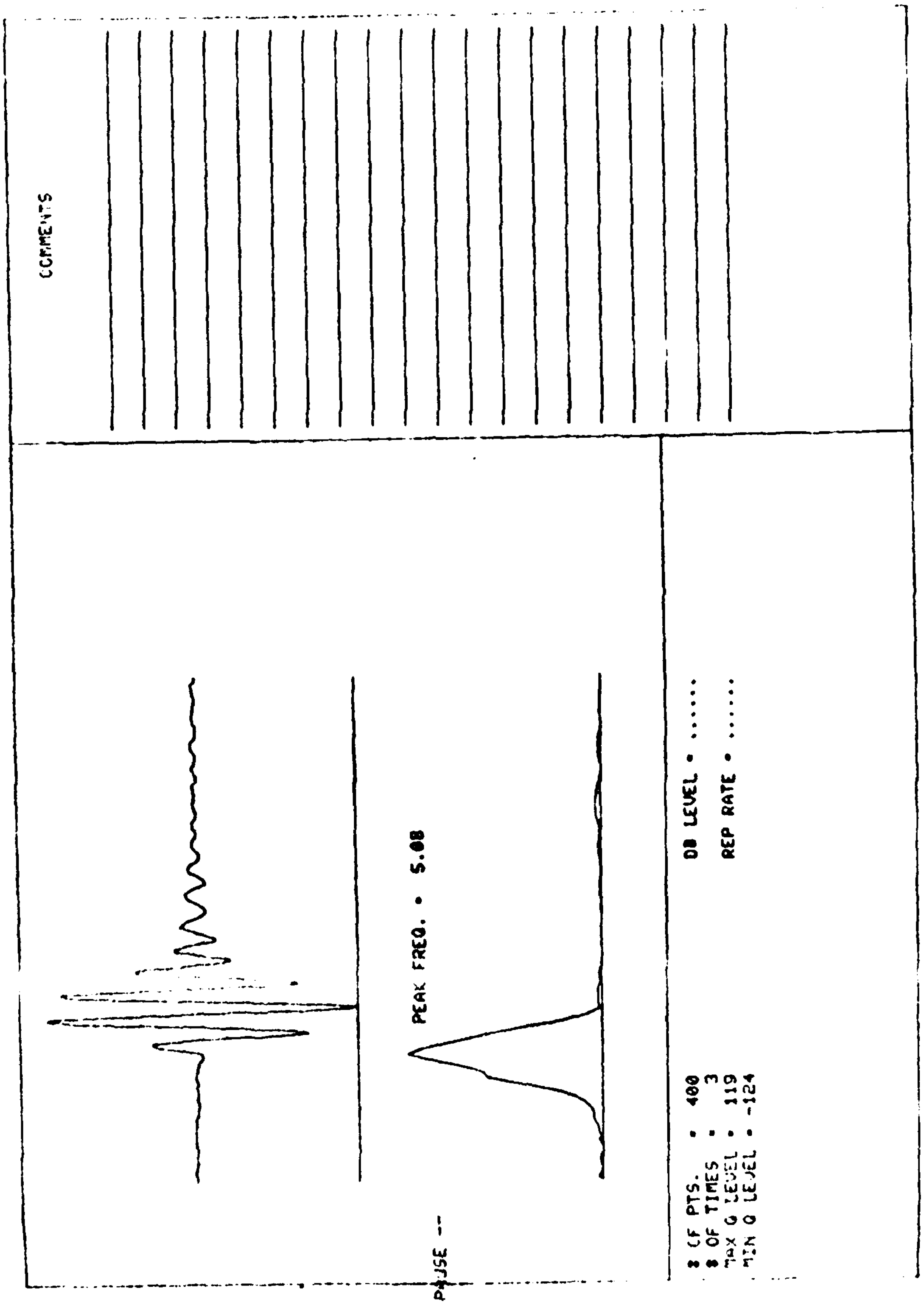


Figure 27. Frequency Frame

Figure 28. Blank Frame

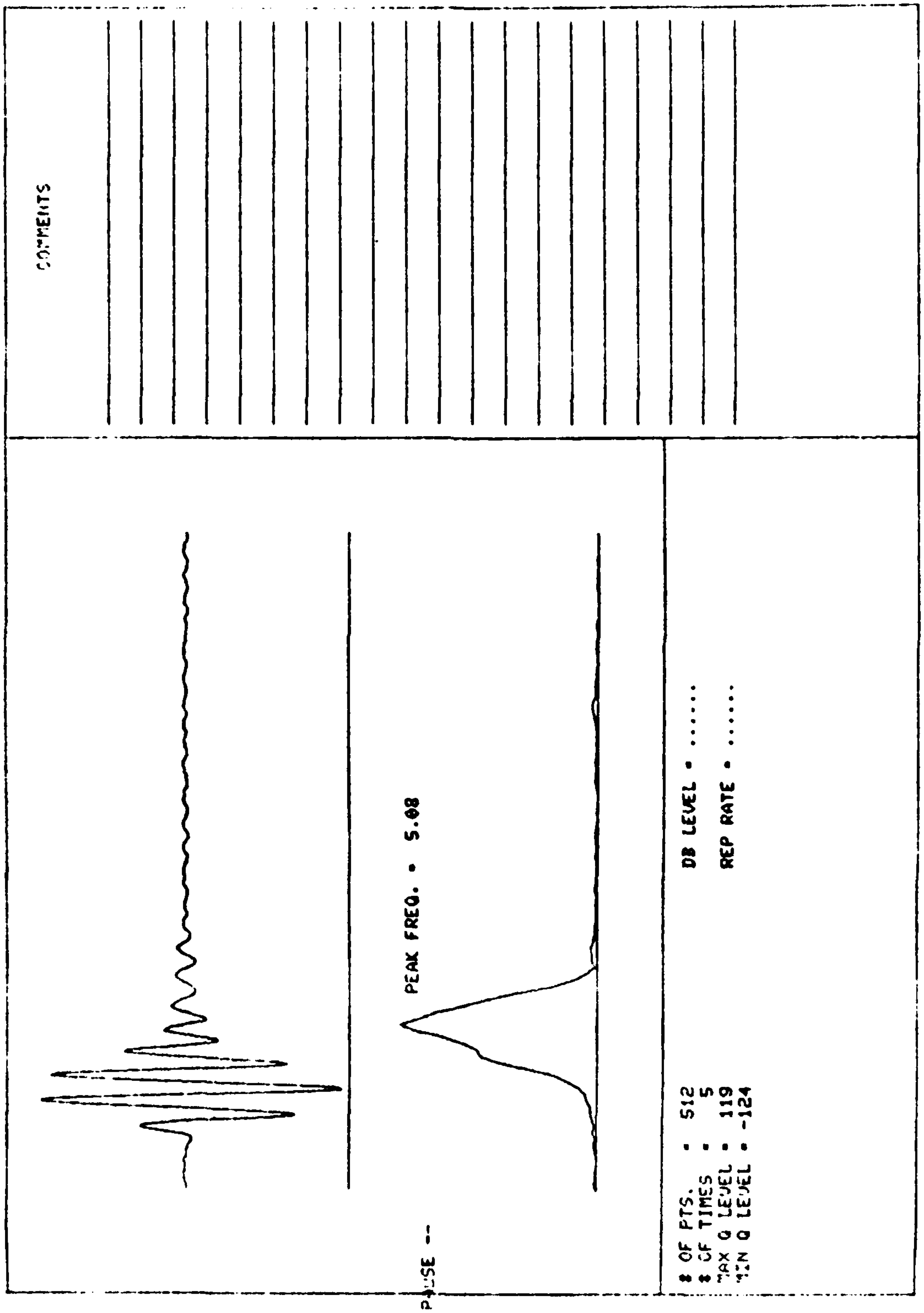


Figure 29. Replay Frame

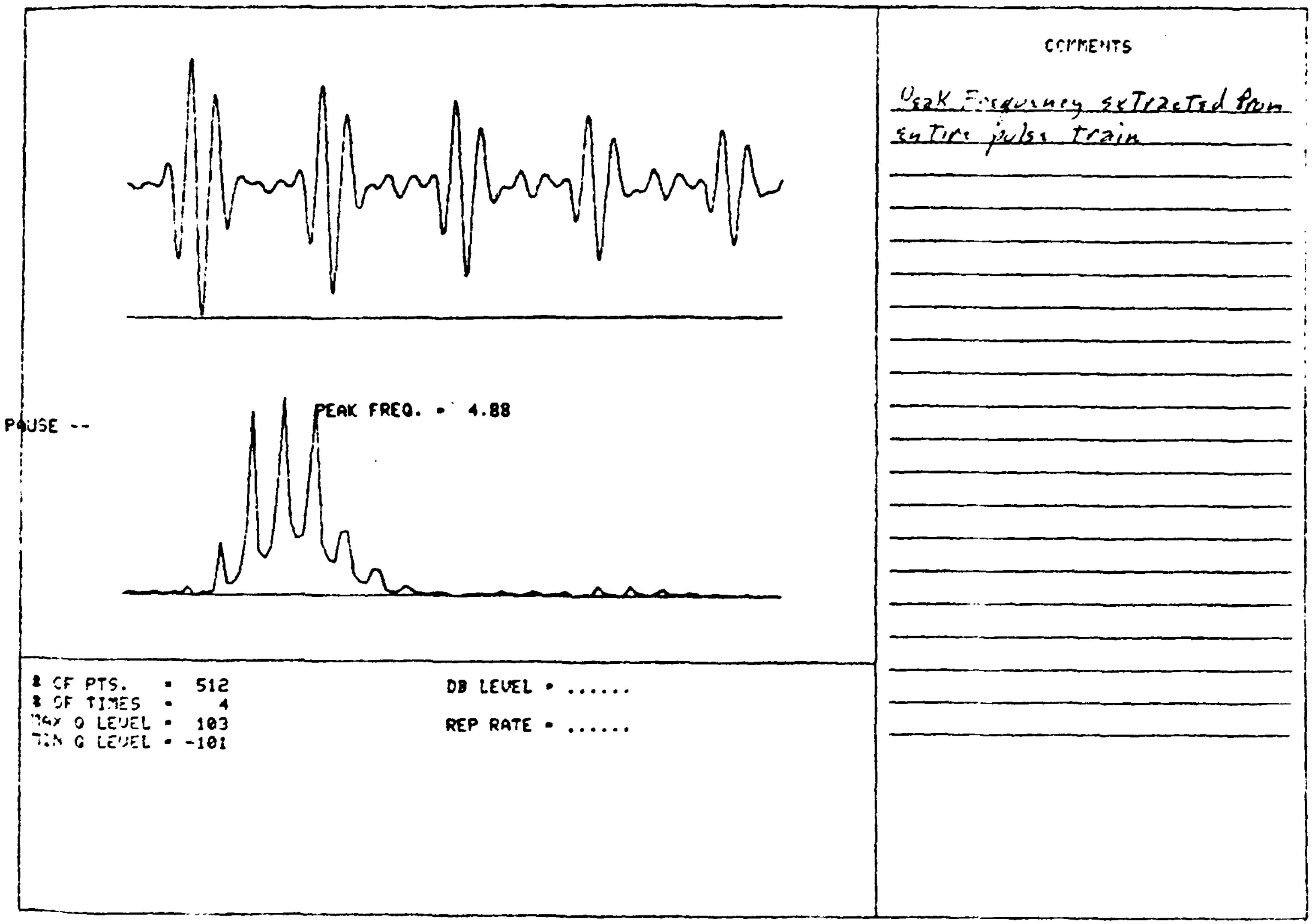


Figure 30. Replay Frame

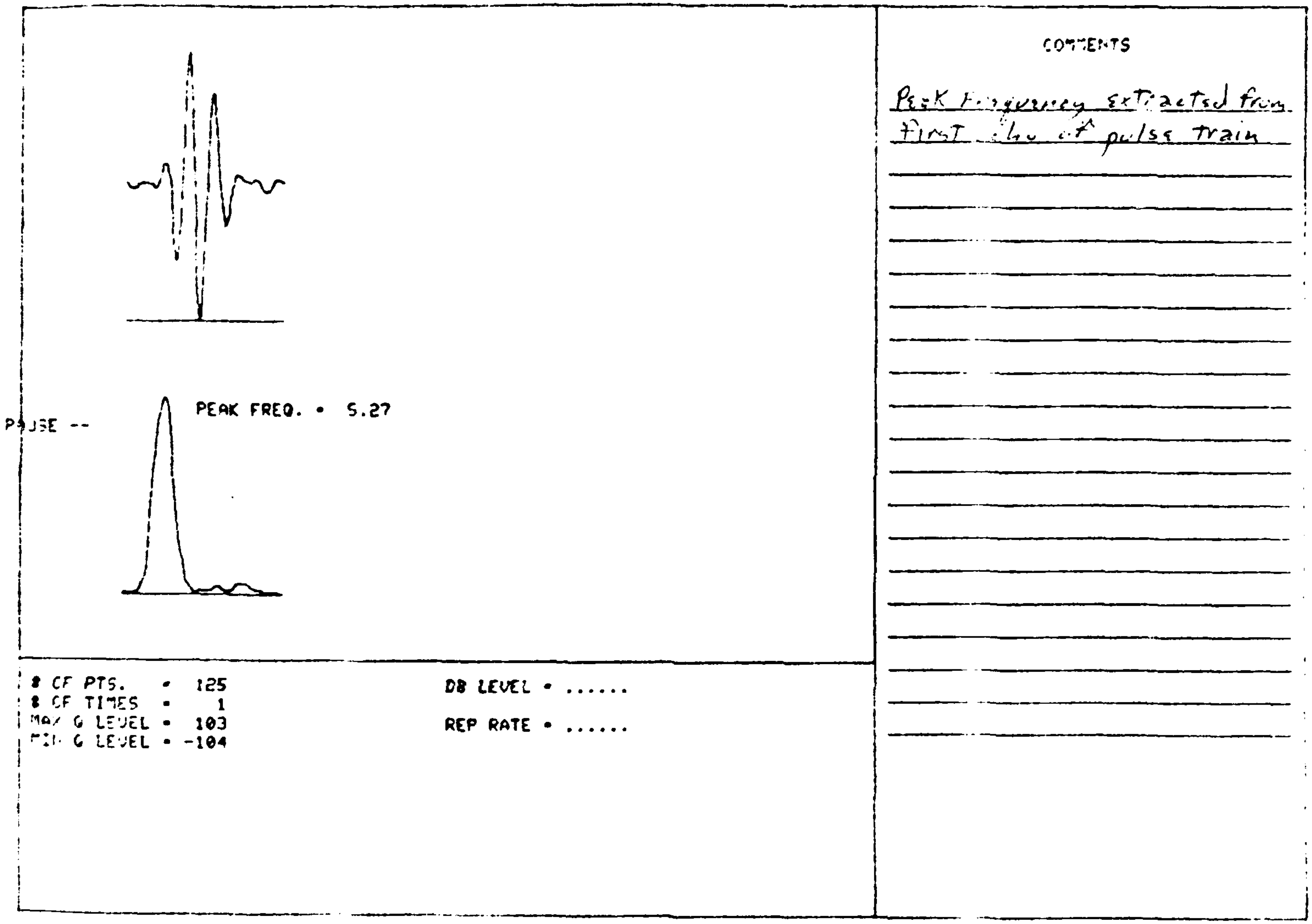


Figure 31. Replay Frame

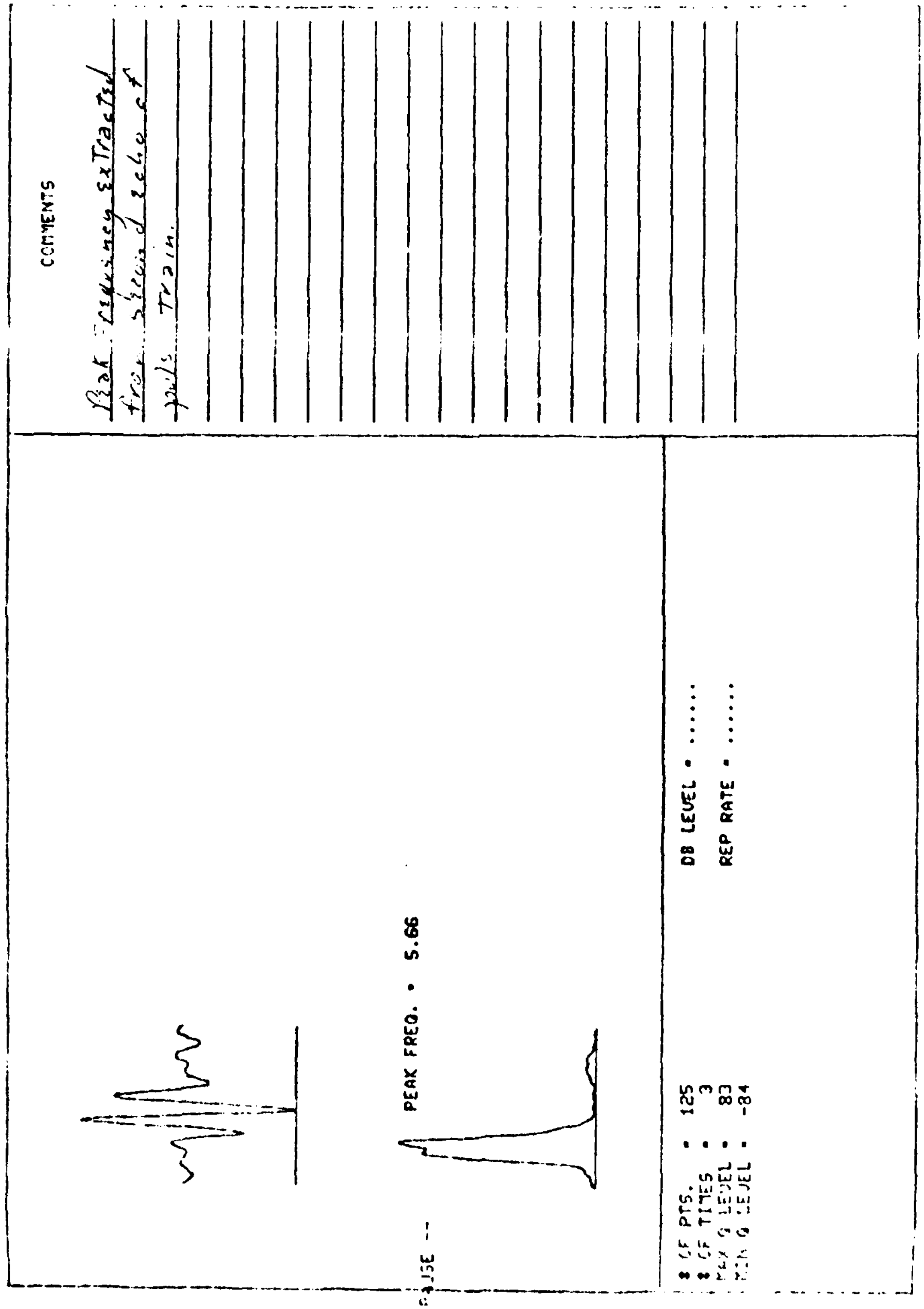


Figure 32. Replay Frame

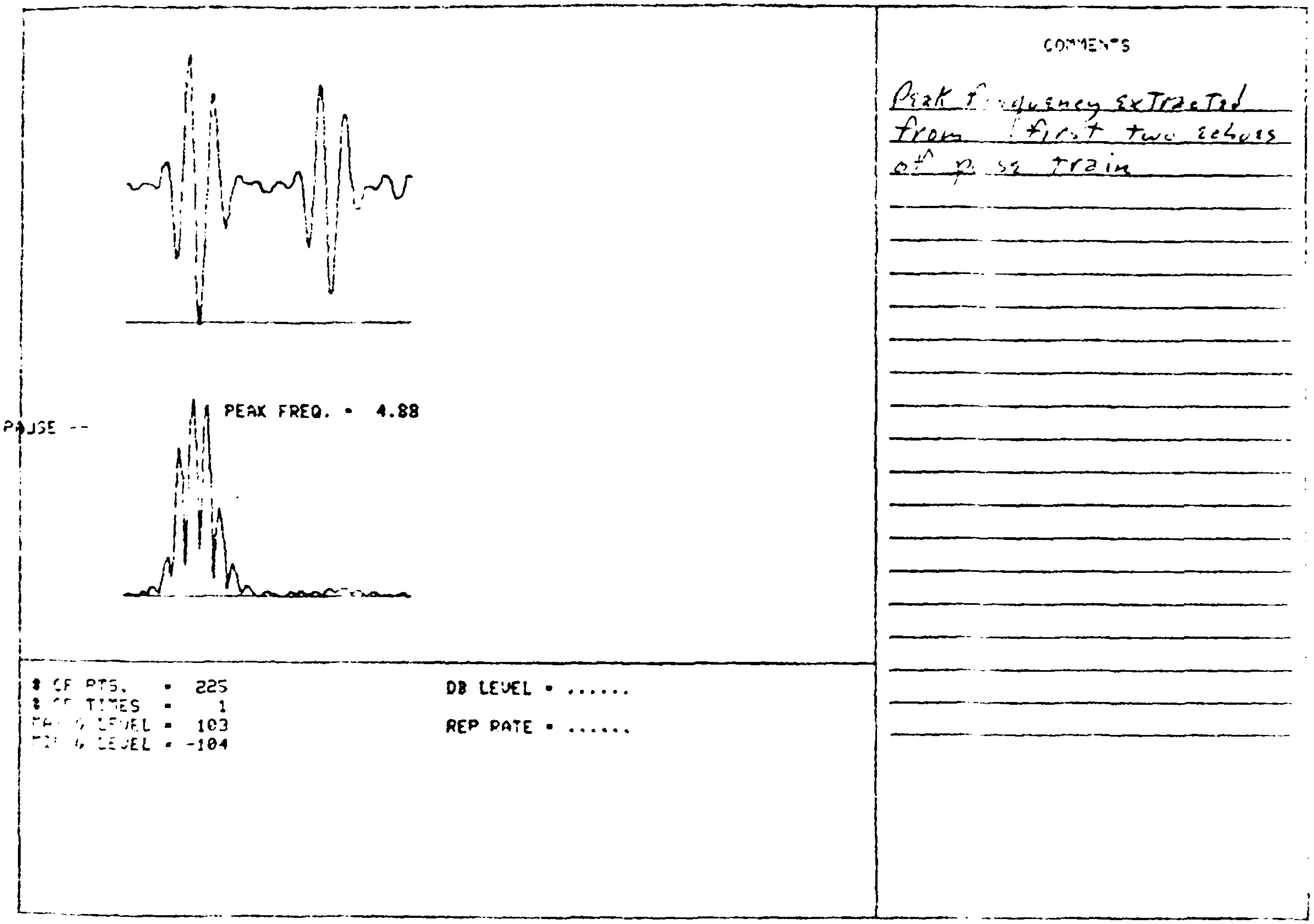


Figure 33. Replay Frame

B. FEATURE EXTRACTION. Feature extraction is the process whereby original and transformed signals are parameterized and the parameters treated as components of a column-like vector called a feature vector. Figure 34 below shows how a waveform might be parameterized.

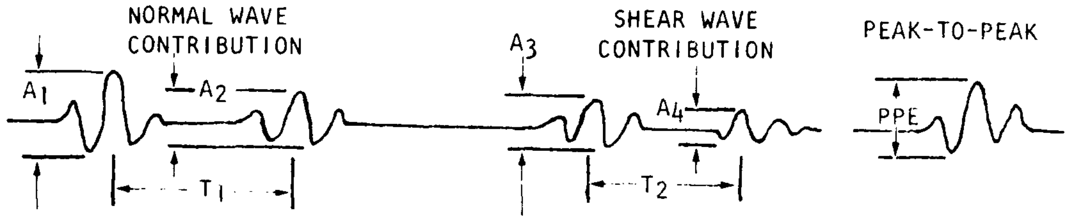


Figure 34. Waveform Parameters

a. Absolute maximum + absolute minimum in either preceding or following half cycle in 0-4μsec window (normal component).
b. Absolute maximum + absolute minimum in either preceding or following half cycle in 8-12μsec window (shear component).
c. Any maxima rising above the 6 dB down (50%) level + absolute minimum in either preceding or following half cycle.

A_1	-	The first peak-to-peak value in time over the 0-4μsec portion of the waveform. A_1 will never be zero.
A_2	-	The second peak-to-peak value in tme over the 0-4μsec portion of the waveform. If there is no second peak-to-peak as defined above, $A_2 = 0$.
A_3	-	The first peak-to-peak value in time over the 8-12μsec portion of the waveform. If no peaks appear in this portion, $A_3 = 0$.
A_4	-	The second peak-to-peak value in time over the 8-12μsec portion of the waveform. If no peaks appear in this portion, $A_4 = 0$.
T_1	-	The time from the maximum value of A_1 to the maximum value of A_2.
T_2	-	The time from the maximum value of A_3 to the maximum value of A_4.

$$\text{feature vector} = \begin{bmatrix} A_1 \\ A_2 \\ A_3 \\ A_4 \\ T_1 \\ T_2 \end{bmatrix}$$

1. All feature extraction is implemented in software. Prior to performing feature extraction, the investigator may wish to examine other representations of the signal. The most popular is its Fourier transform including both the

power spectrum and phase angle. Recently, the so-called cepstrum has been found to be useful in superposition problems. These transforms are available in GANPUI along with the option of arbitrary transforms (operator defined).

2. Some work requires the use of the transfer function of the system under investigation. This is obtained by a process known as Deconvolution. This process may also be used to compensate for changes in transducer characteristics. GANPUI has a deconvolving algorithm and the software to extract features from the transfer function.

3. In those studies where signal amplitudes are hindering rather than helping, appropriate normalizing procedures may be used. Signals may be normalized to have peak-to-peak values of one, to have zero mean with standard deviation of one, and so forth. Transformed data may have to be normalized depending on the nature of the feature extraction involved.

4. This new preprocessed data (transformed, deconvolved, normalized, etc.) is the set of functions on which feature extraction is performed. In general, features are of two broad classes: statistical and physically motivated. Statistical features are those such as mean value, standard deviation, kurtosis, etc. Physically motivated might include parameters such as arrival times, ratios of echo amplitudes to a reference amplitude, spectral depression spacing, etc. There are also two methods of applying these concepts. One is called Global and the other is called Local. Global methods involve the analysis of the entire preprocessed function, whereas Local methods look only at windowed portions of the data. An example of Global and Local feature extraction is shown in Figures 30-33. The feature is the peak frequency of the Fourier spectrum.

C. VECTOR FILING SYSTEM. When ultrasonic inspection is used to predict component performance or quality, the problem is considered either in the form of discrete classes or as a continuum. For example, an adhesive bond strength investigation may classify bonds as good or bad (2 class), high, medium, or low breaking strength (3 class), or continuously by actually predicting the breaking strength in lbs./in.^{2}. Each type of problem requires a different form of algorithm. Since different algorithms require different inputs, feature values (vectors) must be filed in a manner appropriate to the algorithm employed.

1. The general approach is to have three distinct file structures. The first is a file containing only single features over the entire domain of the problem. That is, a sub-file with only feature 1 values, a sub-file with only feature 2 values, etc. Algorithms requiring this type of structure are PDF estimation, 2-space plotting, fuzzy logic, etc. These algorithms will be explained separately.

2. Secondly, a file containing complete vectors is necessary. Algorithms of the ALN (adaptive learning network) type require vector inputs.

3. Lastly, a file structure based on class restricted vectors is required. The Fisher linear discriminant, Minimum Distance, and Factor Analysis algorithms use class restricted vector inputs.

D. PATTERN RECOGNITION PROCEDURES. The first procedure involves the estimation of probability density functions. These curves graphically indicate the probability that a feature will assume a particular value. A narrow unimodal distribution could indicate a poor feature since over all classes, only one value is most likely. The feature would have no merit in differentiating between classes. On the other hand, a multi-modal distribution would indicate a potentially good feature or one that varies hopefully from class to class. See Figure 35.

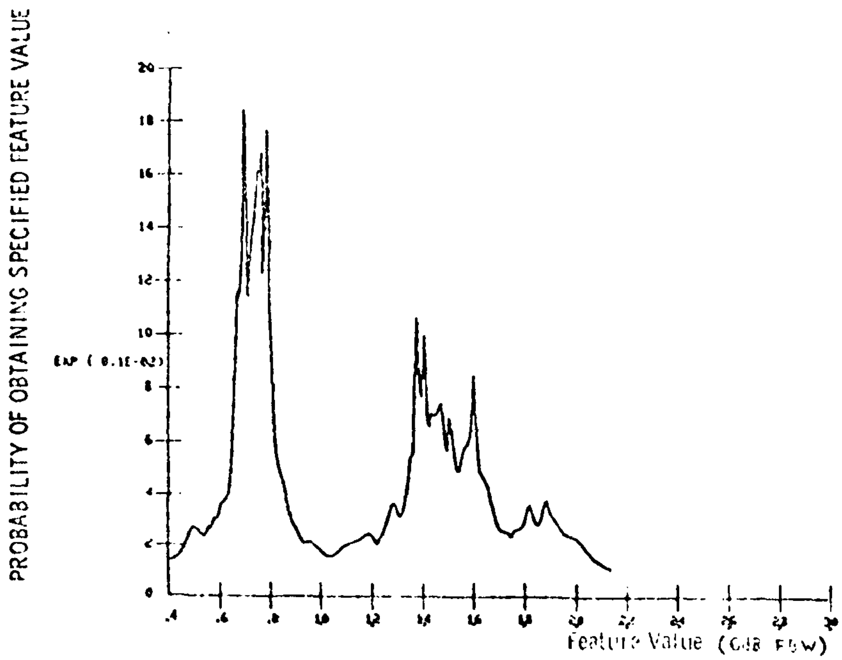


Figure 35. Multi-Modal Distribution

1. Completion of the stage might possibly give insight to the development of a fuzzy type algorithm. This type of reasoning is illustrated in section IVA.

2. A second step is the use of 2-space plots. These are plots of feature I versus feature J for each pair I, J. Plots such as these indicate feature interactions and also have the potential to define distinct clusters, each cluster defining a particular class in the problem framework. Simple algorithms may also be defined on the curves. Examples from a Crack versus Geometry study are shown in Figures 36-38.

3. If simple algorithms are not feasible at this point, more sophisticated algorithms must be used. The results of these initial analyses may be used to give direction to use of the higher level algorithms.

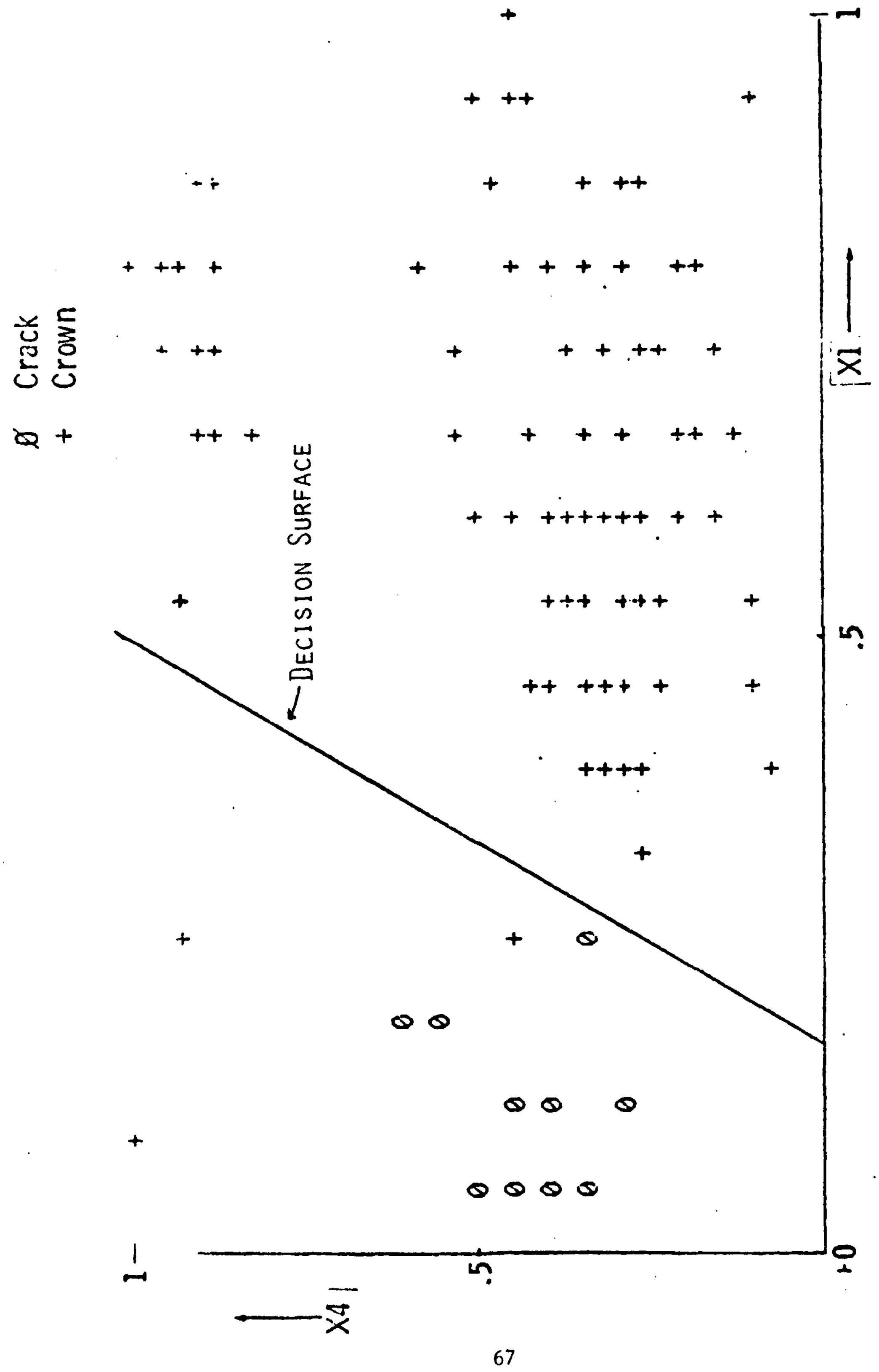


Figure 36. Two Space Plot Between Number of Spectral Depressions Above 20 dB (X1) and 10 dB Down Bandwidth (X4)

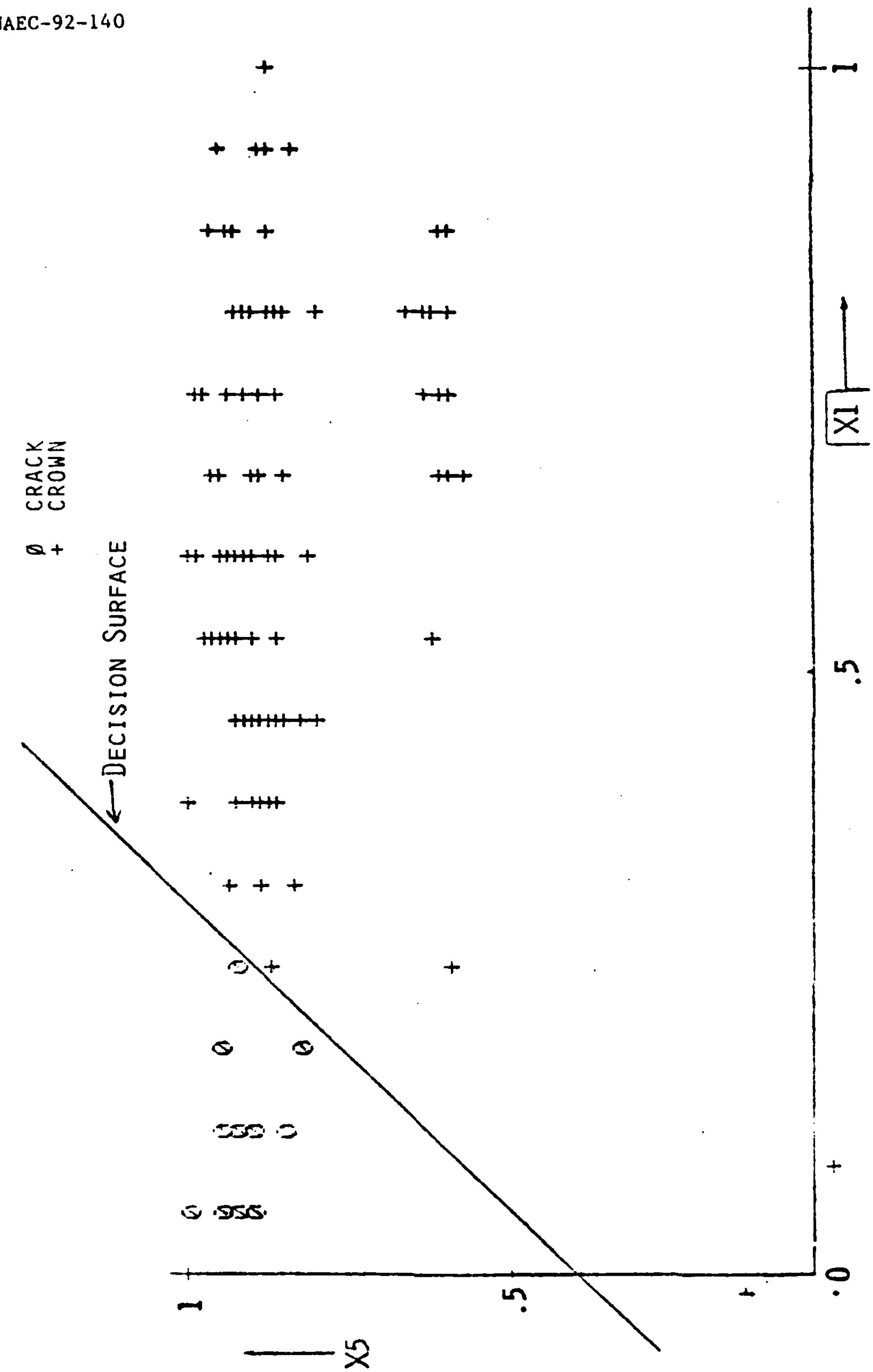


Figure 37. Two Space Plot Between Number of Spectral Depressions Above 20 dB (X1) and 10 dB Down Mid-Frequency (X5)

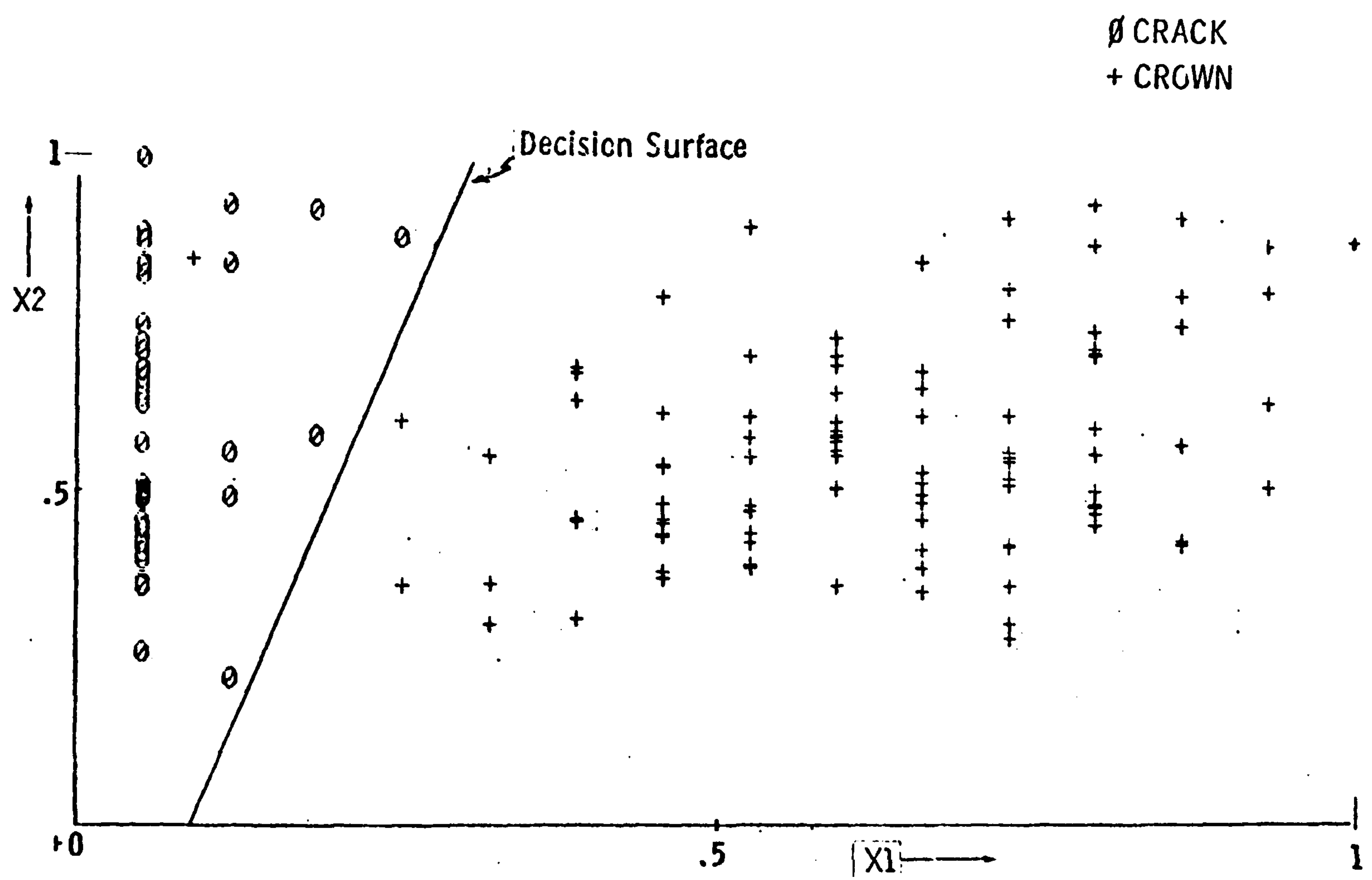


Figure 38. Two Space Plot Between Number of Spectral Depressions Above 20 dB (X1) and Fractional Power Ratio (X2)

4. Before considering complex algorithms, PDF curves may be used again. This time the curves are plotted on a feature by class basis. An example of a typical plot is shown in Figure 39. The shaded regions indicate where values on the x-axis would indicate a high bond. Other regions are most likely to contain values obtained from low-strength bonds. An algorithm may possibly be implemented at this stage.

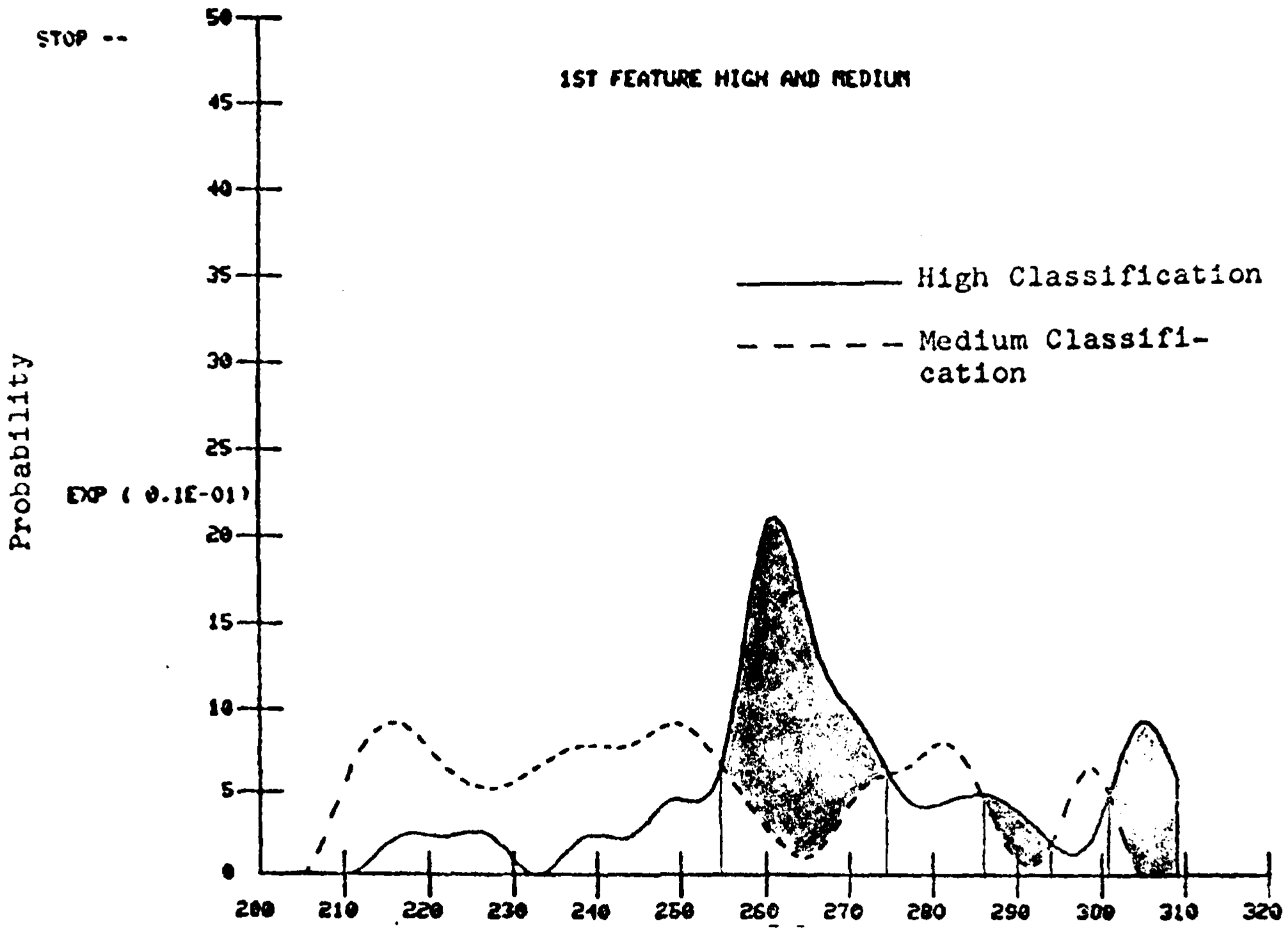


Figure 39. Use of PDF Curves

5. Factor Analysis is used next. This is a statistical procedure for determining those features which contribute most to variation in classes. It is a method for ordering features with respect to their degree of importance in the class discrimination problem. See the section on Factor Analysis for a detailed description of this method.

6. The above procedures essentially define those features that will be useful. Using this set of "good" features, three other techniques are possible.

7. A minimum distance classifier is an algorithm requiring two sets of statistically similar data. One set is used for training and the other for evaluation of the algorithm. The first set is called the prototype set. The evaluation set is called the test set. Training information is used in establishing the prototype vectors along with the minimum distance classification routine. The procedure works quite simply by examining a distance in an "n' dimensional space, "n" being the number of elements in the feature vector. A test vector is compared to the two prototype vectors by a distance formula. The test vector is classified according to the resulting distance measure which classifies according to the prototype it is closest to. The formulas used are summarized below.

$$d_{T1} = \sqrt{(X_{T1} - X_{P11})^2 + (X_{T2} - X_{P12})^2 + \ldots + (X_{Tn} - X_{P1n})^2}$$

$$d_{T2} = \sqrt{(X_{T1} - X_{P21})^2 + (X_{T2} - X_{P22})^2 + \ldots + (X_{Tn} - X_{P2n})^2}$$

which reduces to

$$d_{TN} = \sqrt{\sum_{i=1}^{n} (X_{Ti} - X_{PNn})^2}$$

N = prototype number

n = number of elements in the feature vector

d_{TN} = distance between the test vector and the prototype vector N

if $d_{T1} < d_{T2}$ then class 1

if $d_{T2} < d_{T1}$ then class 2

VI. SAMPLE PROBLEMS USING GANPUI

A. GENERAL CONSIDERATIONS. This section will follow the paths through GANPUI that have led to the development of two successful algorithms. The problems to be discussed are intergranular stress corrosion cracking and adhesive bonds.

1. The first major step after problem definition is transducer selection. This involves specification of frequency ranges, mode of usage, immersion, contact, boot, etc., the choice of narrow band or broad band, and whether or not a single or dual element probe should be used. The inherent noise levels of the problem are a major factor in the choice of these parameters. Low noise levels facilitate the use of transducer compensation routines. The transfer function of the system may also be used as a source when low noise levels are involved. Higher feature levels indicate that tight specifications are necessary for transducers that are different from the design transducer.

2. The question arises, will features be transducer dependent or independent? The answers to this type of question establish the data acquisition protocol.

3. The next step depends on the physics and mechanics of the problem. This is the choice of feature sources and those features that are to be extracted. Questions like, is superposition possible?, are there frequency shifts involved?, what are possible attenuation effects?, etc., are all indicative of the particular features that are required.

4. Finally, using the acquired feature vectors, pattern recognition methods are tried. The progress is from simple to sophisticated, simpler solutions being favored. Once several algorithms have been attempted, the trade-offs between simplicity, reliability, and economy are evaluated. Then, in a sense, the optimum scheme is implemented. See Appendices and references 1, 2, and 3 below for further details.

B. SAMPLE PROBLEMS.

	Bonds	Cracks
Mode	immersion	contact
Frequency Range	10 MHz	1.5-3 MHz
Type	Single element Broad band	Dual element Narrow band
Signal Processing	Spatial Averaging Signal Averaging	Signal Averaging ----
Noise Level	Low Deconvolution	High ---
Feature Sources	Transfer function	Video envelope Fourier spectrum
Features		Pulse Duration Partial energy in spectrum
Algorithm	Fisher linear discriminant	Two space plot
Performance (# correct/total)	---	---

C. REFERENCES

1. Rose, Joseph L. and Avioli, Michael J., "Transducer Compensation Concepts in Flaw Classification," presented at the Spring ASNT meeting in New Orleans, 1978, and to be published in a reliability supplement to Materials Evaluation.

2. Rose, Joseph L. and Singh, Gurvinder P., "An Ultrasonic Waveform Feasibility Study - Crack (IGSCC) Vs. Geometric Reflector (Crown) Classification In 304 Stainless Steel Pipe Weld Specimens," to be presented in the British Journal of Nondestructive Testing.

3. Rose, Joseph L. and Thomas, Graham H., "The Fisher Linear Discriminant Function for Adhesive Bond Strength Prediction," to be published in the British Journal of Nondestructive Testing.

This page left blank
intentionally.

APPENDIX A

TABLE A-1. INPUT DATA TO FACTOR ANALYSIS PROGRAM

CASE	FEATURE 1	FEATURE 2	FEATURE 3	FEATURE 4	FEATURE 5
1	1	.052	7.071	.781	1.465
2	1	.086	7.387	.879	1.514
3	1	.053	7.552	.879	1.514
4	2	.088	7.319	.781	1.562
5	1	.029	6.780	.684	1.514
6	7	.082	12.401	.781	1.514
7	6	.071	10.103	.537	1.440
8	11	.079	12.786	1.660	.879
9	13	.045	11.900	.684	1.514
10	10	.049	9.460	.684	1.367

TABLE A-2. MEANS AND STANDARD DEVIATIONS

FEATURE	MEAN	STANDARD DEVIATION
1	5.100	4.932
2	.062	.019
3	.823	2.509
4	.923	.402
5	1.360	.258

This page left blank
intentionally.

APPENDIX B - THE UTILITY OF GRAPHICAL AIDS

A. PATTERN RECOGNITION. Pattern recognition studies often involve the use of spaces with high dimensionality. A feature vector of dimension six is not uncommon. One way to graphically display such a vector is through the use of closed polygons of equal sides. This, of course, applies to vectors of dimension three or greater, an equilateral triangle being the closed polygon with the least number of sides. Consider the polygon shown in Example B-1. Each side is length L. Each side may be considered the range of a vector component, if the component were normalized to span the interval 0, L . This is easily accomplished by applying the following formulation.

Let a = the minimum value a feature can assume
Let b = the maximum value a feature can assume
Let L = the desired range to be spanned by the feature

Then $F(f) = \frac{f - a}{b - a} \cdot L$ will map feature values f onto the range 0, L .

Consider the set of features f_1, f_2, and f_3 with corresponding ranges

$-10 \quad f_1 \quad 5$
$15 \quad f_2 \quad 20$
$-50 \quad f_3 \quad 75$

A typical feature vector might be

$$\begin{bmatrix} -2.5 \\ 18.75 \\ 70 \end{bmatrix}$$

Assume it is desired to visualize this vector via a triangle with sides 10 units long.

For f_1, $a = -10$, $b = 5$ $\qquad -2.5 \qquad \frac{-2.5 + 10}{15} \cdot 10 = 5 = f_1^T$

For f_2, $a = 15$, $b = 20$ $\qquad 18.75 \qquad \frac{18.75 - 15}{5} \cdot 10 = 7.5 = f_2^T$

For f_3, $a = -50$, $b = 75$ $\qquad 70 \qquad \frac{70 + 50}{125} \cdot 10 = 9.6 = f_3^T$

The graphical representation of $\begin{bmatrix} -2.5 \\ 18.75 \\ 70 \end{bmatrix}$ is shown in Example B-2.

The shaded region is the pattern of interest. Now consider the vector

$$\begin{bmatrix} 0 \\ 19 \\ 20 \end{bmatrix} \qquad \begin{bmatrix} 6.67 \\ 8 \\ 2.4 \end{bmatrix}^T$$

This is shown in Example B-3.

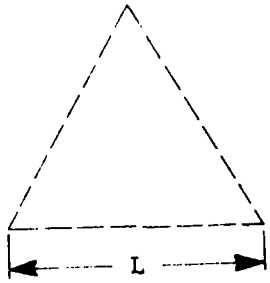
1. It is noted that a different pattern is generated for a different vector. Also, for vectors that are relatively "close" or clustered, the generated patterns are similar; vectors that are declustered (not in one particular cluster but in another) generate different patterns.

2. The efficient use of this display method comes when dealing with higher dimension vectors. Shown in Example B-4 is an example for a six-dimensional vector.

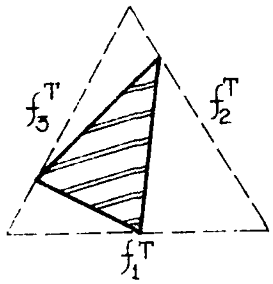
3. Parametric plotting may also be useful as an aid to pattern recognition studies. Illustration of this technique is best shown by example. Three Fourier Spectrums are shown in Example B-5. The one on the right may be considered as a "reference" spectrum. The remaining two spectrums may be assumed to be derived from two different classes of time functions. At each frequency there corresponds a value on the reference spectrum, $r(f)$, and values on the other two spectrums, $s_1(f)$ and $s_2(f)$ respectively. Using the x-axis as a $r(f)$ axis and the y-axis for $r(f)$, $s_1(f)$, and $s_2(f)$ axes, the parametric plots shown in Example B-6 are obtained.

4. Another technique is one involving the use of so-called "pie" graphs. This method is generally applied to curves whose areas may be normalized to one. Fourier spectrums are good examples. See Example B-7a.

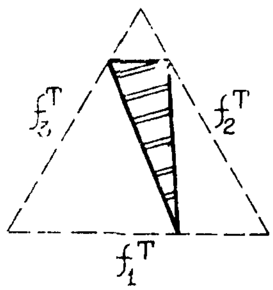
5. This unit area can then be related to a circle or "pie" of unit area, Example B-7b. The original curve can be sectioned into intervals or bands, each of equal length. See Example B-8a. The intervals will contain certain percentages of the total area. These percentages correspond to "slices" of varying size within the "pie". Example B-8b illustrates this concept. Instead of denoting circle sectors by numbers, a gray scale code or symbolic patterns may be used. Example B-9 shows a typical display mode.



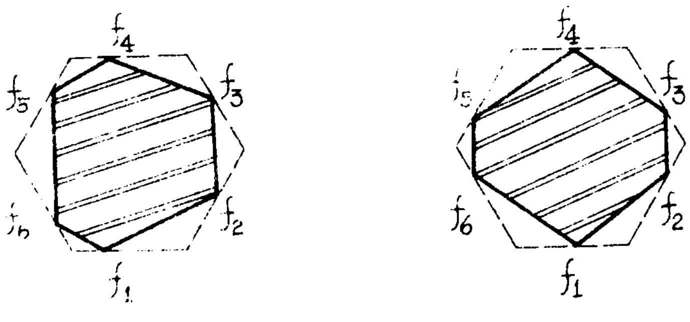
Example B-1



Example B-2

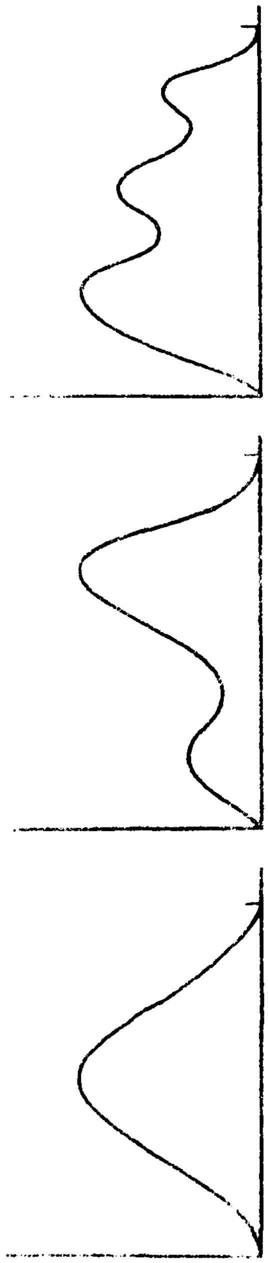


Example B-3

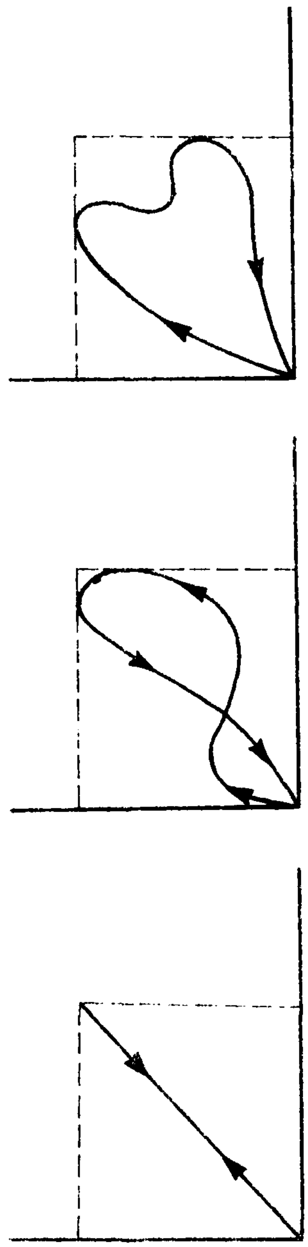


Example B-4

epresentation of N-Dimensional Vectors

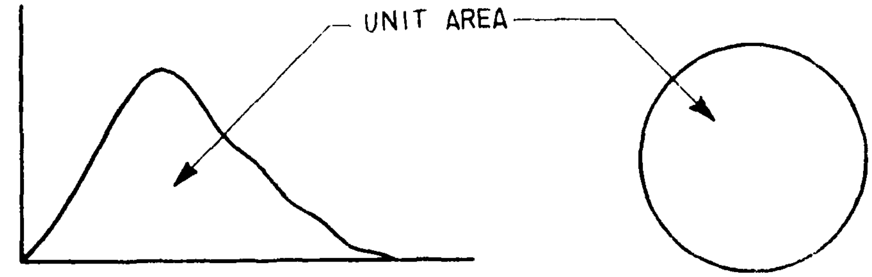


Example B-5



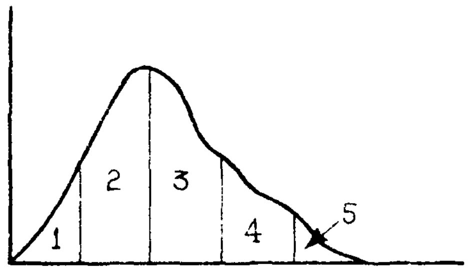
Example B-6

Examples of Parametric Plotting

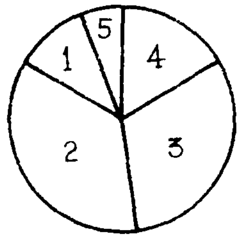


Example B-7a

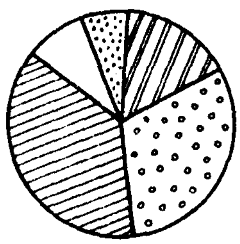
Example B-7b



Example B-8a



Example B-8b



Example B-9

Examples of Pie Graphing

This page left blank
intentionally.

APPENDIX C - (ABSTRACT) A PATTERN RECOGNITION REFLECTOR CLASSIFICATION FEASIBILITY STUDY - CRACK (IGSCC) VS. GEOMETRIC (CROWN) REFLECTOR IN 304 STAINLESS STEEL PIPE WELD SPECIMENS

A feasibility study has been conducted in order to evaluate the potential of pattern recognition techniques for discriminating between geometrical and crack reflector signals obtained during ultrasonic inspection of the weld zone in 304 austenitic stainless steel pipes. A geometrical reflector is defined as a reflector associated with the weld geometry and/or a flaw incapable of causing catastrophic failure e.g. crown, counterbore, suck-back, drop-thru, etc. Seven welds from four different 4" diameter pipe specimens, containing intergranular stress corrosion cracking (IGSCC) were examined ultrasonically. The ultrasonic inspection was conducted in a pulse echo mode using a 1.5 MHz nominal center frequency, 3/8" diameter transducer mounted on a plexiglass shoe with 45° refracted transverse wave insonifying the area of interest. The ultrasonic data was correlated with the dye penetrant tests and ultrasonic examination conducted by Southwest Research Institute (SWRI) in order to obtain valid training information. The data in this particular feasibility study included crown geometric reflectors and crack reflectors. A total of 107 crown indications and 40 intergranular stress corrosion cracking indications were analyzed. The analysis did not consider any arrival time, amplitude information or, in fact, any other time domain features, but was based on various Fourier transform features. A 100% reliability level was obtained for discriminating an IGSCC indication vs. crown indication using automated pattern recognition algoritnm.

The overwhelming success of the pattern recognition algorithm employed in this study demonstrates the applicability of this technique for solving such important problems as discrimination between IGSCC vs. geometric reflectors in 304 stainless steel pipe welds. Work on other kinds of geometric reflectors is in progress for establishing an overall reliability level in reflector classification analysis.

This page left blank
intentionally.

APPENDIX D - (ABSTRACT) "TRANSDUCER COMPENSATION CONCEPTS IN FLAW CLASSIFICATION"

by
Joseph L. Rose, Professor of Mechanical Engineering
and
Michael J. Avioli, Graduate Student, Mechanical Engineering and Mechanics

Flaw classification analysis is quite often strongly influenced by the type of ultrasonic waveform that is generated by an ultrasonic transducer. One goal of this paper is to introduce procedures that could possibly make flaw classification algorithms become somewhat independent of certain ultrasonic waveform characteristics being used in the data acquisition procedure. Data acquisition of ultrasonic pulse echo signals depends quite strongly on many test system characteristics, in particular, special characteristics of the ultrasonic transducer and pulser-receiver instrument characteristics. A transducer compensation procedure is presented in this work that requires a suitable reference signal containing "noise" contributed only by system components external to the unknown flaw, and in software, a processing scheme is designed to remove external effects, therefore, allowing concentration on flaw characteristics contained within the ultrasonic signal.

The processing scheme has four general components: Acceptor, Compensator, Comparator, and Evaluator. The acceptor is basically a gate that decides whether or not a particular transducer is usable for the problem at hand. The compensator implements a mathematical deconvolution process. The comparator does a feature by feature similarity check on the desired signal and the compensated signal. The evaluator is any scheme that can determine the performance of a given transducer. In particular, the algorithm under study may be used to evaluate transducer performance. The evaluation stage is followed by an examination of the comparator results. Tolerances relating to the acceptability of a transducer are obtained through this final stage.

Model analysis is used to study the compensation problem. A Layered Model is used with various levels of system "noise" being introduced, in order to examine the "noise" effects in the deconvolution computation process. Promise for attaining success in this difficult compensation problem is good, particularly when considering signal averaging as a signal processing tool.

This page left blank
intentionally.

APPENDIX E - (ABSTRACT) THE FISHER LINEAR DISCRIMINANT FUNCTION FOR ADHESIVE BOND STRENGTH PREDICTION

An ultrasonic inspection system for the prediction of adhesive bond strength for metal-to-metal applications is of great value to many government and industrial agencies throughout the world. The prediction of adhesive bond strength based on surface preparation, assuming that there are no delaminations, inclusions, or such cohesive type problems as improper curing, etc. is the goal of this study. Ultrasonically evaluating adhesive bonds that have partially delaminated, is generally easily accomplished by using C-scan techniques, but a major problem arises when the deficiency in the bond is either adhesive or cohesive in nature. Our study involved primarily the adhesive aspect of the bond strength, which is related to the surface preparation problem. Test specimens were manufactured so that an improper surface preparation occurred on either or both substrates in an aluminum-to-aluminum step-lap joint. The specimens with little or no surface preparation provided weak bonds and the specimens with proper surface preparations, in general, produced strong bonds.

A resource base developed in earlier years in experimental technology, theoretical ultrasonic wave interaction studies with adhesive bond models, manufacturing technology, and shear stress distribution analysis have all been incorporated into a pattern recognition program of study. Such topics as nearest neighbor philosophy, fuzzy logic analysis, probability density function analysis, and adaptive search and learning techniques for linear and non-linear models have been investigated. A Fisher Linear Discriminant algorithm has been developed which affords a 91% reliable prediction for adhesive bond strength. Unfortunately, results indicate that the prediction algorithms depend strongly on the particular transducer which was initially used for data acquisition. Data acquired with a second transducer, having different pulse form characteristics, in general, did not provide reliable results for predicting adhesive bond strength. To compensate for the transducer differences, a deconvolution technique was implemented to expand the selection of useful transducers. Limited success on this technique has been obtained to date because of the inherent system noise in ultrasonic data acquisition equipment.

A completely automated ultrasonic inspection system has been developed for predicting bond strength in metal-to-metal adhesively bonded step-lap joints. Results to date provide a 91% reliability for solving this difficult problem of predicting adhesive bond performance.

This page left blank
intentionally.

APPENDIX F - ON THE UTILITY OF PROBABILITY DENSITY FUNCTION ANALYSIS

Probability density function curves can be useful for solving a large number of engineering problems. Of primary concern, of course, in pattern recognition, probability density functions can be used for feature selection. Pattern recognition problems call for the establishement of a feature vector that can be used for developing a reflector classification algorithm. In addition to this very important application of PDF curves for feature evaluation, a variety of other applications in engineering is being considered today. The principles of probability density function analysis are particularly suited to an inspection philosophy for composite materials. A brief review of possible applications is outlined below.

1. To evaluate Material Uniformity in a Quality Control Test - Composite materials, because of the dual material content, variation in fabrication, and an isotropic character, etc., are noisy with respect to ultrasonic waveform content as reflected from the composite structure. A good composite material will have a PDF curve for a particular feature that is fairly tight. Experimental analysis, of course, can acquire this PDF information or "PDF signature". Uniformity of the composite material can, therefore, be evaluated since poorly manufactured composites would have a different PDF signature and most probably producing that of a wider and distorted PDF curve. A material acceptance criteria could, therefore, be written as a function of tolerances on the PDF curves.

2. Material Selection Philosophy for Improved Inspectability - Quite often a number of fabrication techniques are considered in the development of a composite material. As strength, temperature, and moisture tests indicate that the performance of the composite is independent of the fabrication process, it is proposed to select the fabrication process with the greatest inspectability. Inspectability of the material can be established on the basis of PDF signatures for the composite materials and their corresponding fabrication process. Again, a tightly grouped PDF curve for certain features is highly desirable.

3. Material Lay Up Selection - Composite materials can be laid up at a variety of fiber orientations and lay ups. In some cases, the angle ply lay up procedures are important for improved composite material performance. On the other hand, performance may not be improved. In this case, it is suggested that PDF signatures for the various composite material angle ply configurations be used to select the lay up that is most respectable, again following some of the logic developed earlier for tightly grouped PDF curves.

4. Transducer Selection - Probability density function curves can even be used to select transducers for material inspection. As an example a comparison of single element versus dual element transducer application for a composite material can be evaluated. Dual element transducer work will have composite material inspection because of the removal of back scatter ultrasonic radiation. This physical principle of recording only forward scatter information can be demonstrated quite nicely by examining a number of features in a probability density function analysis. Nicely distributed and tight PDF curves could be used to select the best transducer for a particular application.

5. Damage Evaluation - It is quite obvious that once a PDF signature is acquired for a composite material that a PDF curve that is produced at some later date that indicates some marked change is indicative of material change or degradation. In most cases, the change would come about because of crack, delamination, or environmental degradation. PDF curves produced at various areas of a composite material can, therefore, be used to provide us with damage propagation information.

APPENDIX G - SELECTED READINGS

SIGNAL PROCESSING AND PATTERN RECOGNITION

1. Nilsson, Nils J., Learning Machines, McGraw-Hill Book Company, New York, New York, 1965.

2. Hsu, Hwei P., Fourier Analysis, Simon & Schuster, Inc., New York, New York, 1970.

3. Sammon, John W., Jr., "Interaction Pattern Recognition Analysis and Classification," IEEE Transactions on Computers, Vol. C-19, No. 7, July 1970.

4. Maisel, Louis, Probability, Statistics and Random Processes, Simon & Schuster, Inc., New York, New York, 1971.

5. Kemerait, R. C., Childers, D. G., "Signal Detection and Extraction by Cepstrum Techniques," IEEE Transactions on Informations Theory, Vol. IT-18, No. 6, November 1972.

6. Chen, C. H., Statistical Pattern Recognition, Hayden Book Company, Inc., Rochelle, New Jersey, 1973.

7. Duda, Richard O., Hart, Peter E., Pattern Classification and Scene Analysis, John Wiley & Sons, Inc., New York, New York, 1973.

8. Seydel, James A., Frederick, Julian R., "A Computer-Processed Ultrasonic Pulse Echo System," Materials Evaluation, Vol. XXXI, No. 11, November 1973.

9. Bolch, Ben W., Huang, Cliff J., Multivariate Statistical Methods for Business and Economics, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.

10. Brigham, E. Oran, The Fast Fourier Transform, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.

11. Sanyal, Probal K., Foley, Donald H., Evans, Edward A., "An Interactive Waveform Processing System (WPS)," PAR report #74-21, Pattern Analysis & Recognition Corp., Rome, New York, October 1974.

12. Mucciardi, Anthony N., et al. "Adaptive Nonlinear Signal Processing for Characterization of Ultrasonic NDE Waveforms," Task I, Inferences of Flat-Bottom Hole Size, USAF Materials Laboratory, Air Force Systems Command, Wright-Patterson AFB, OH, by Adaptronics, Inc., McLean, VA, January 1975.

13. Schwartz, Mischa, Shaw, Leonard, Signal Processing, McGraw-Hill Book Company, New York, New York, 1975.

14. Hamming, R. W., Digital Filters, Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1977.

15. Harris, Fredric J., "On the Use of Windows for Harmonic Analysis with the Discrete Fourier Transform," Proceedings of the IEEE, Vol. 66, No. 1, January 1978.

16. Nabel, E., Mundry, E., "Evaluation of Echoes in Ultrasonic Testing by Deconvolution," Materials Evaluation, Volume 36, No. 1, January 1978.

17. Rose, Joseph L., Avioli, Michael J., "Transducer Compensation Concepts in Flaw Classification," presented at the Spring ASNT meeting in New Orleans, 1978, and to be published in a reliability supplement to Materials Evaluation.

AD-A083 931 DREXEL UNIV PHILADELPHIA PA DEPT OF MECHANICAL ENGIN--ETC F/G 14/2
GENERALIZED APPROACH TO NEW PROBLEMS IN ULTRASONIC INSPECTION (--ETC(U)
APR 80 J L ROSE, M J AVIOLI N68335-78-C-1041
UNCLASSIFIED NAEC-92-140 NL

2 OF 2
AD
A083931

END
DATE
FILMED
-80
DTIC

ACKNOWLEDGMENT

Krautkramer-Branson, Inc., is acknowledged for their computer support and also for their support of many of the topics on pattern recognition that are included in this report.

DISTRIBUTION LIST

NAVAIRENGCEN
1115
9011
92A1B
92A12 (2)
92713
92724 (20)

NAVAIRDEVCEN 60633 (2)
NAVAIRPROCEN PE43
AFML/LLP
AMMRC DRXMR-MQ
NAVSHIP R&D CENTER 2803

NAVAIRSYSCOM
AIR-340E (3)
AIR-5523 (2)
AIR-4114 (2)
AIR-5203 (2)
AIR-950D (2)

DDC (12)

REVISION LIST

REVISION	PAGES AFFECTED	DATE OF REVISION